ELEMENTARY STATISTICS

Henry E. Garrett, Ph.D.

Professor of Psychology
Columbia University

LONGMANS, GREEN AND COMPANY
NEW YORK · LONDON · TORONTO · 1956

LONGMANS, GREEN AND CO., INC.
55 FIFTH AVENUE, NEW YORK 3

LONGMANS, GREEN AND CO., LTD.
6 & 7 CLIFFORD STREET, LONDON W 1

LONGMANS, GREEN AND CO.
20 CRANFIELD ROAD, TORONTO 16

ELEMENTARY STATISTICS

PUBLISHED SIMULTANEOUSLY IN THE DOMINION OF CANADA BY LONGMANS, GREEN AND CO., TORONTO

FIRST EDITION

LIBRARY OF CONGRESS CATALOG CARD NUMBER 56-6220

PRINTED IN THE UNITED STATES OF AMERICA

VAN REES PRESS • NEW YORK

PREFACE

This little book and its accompanying workbook have been written to provide an introduction to statistical method for students in psychology and in the social sciences. The first five chapters are concerned with descriptive statistics: the treatment of the frequency distribution and its graphical representation. Chapters 6 and 7 outline in simple fashion the role of the normal probability curve in mental measurement and the problem of testing experimental hypotheses, i.e., making inferences from sample to population. Chapters 8, 9, and 10 deal with rank order and linear correlation, and with two frequently useful topics, namely, the Chi-square test and methods of comparing and combining scores.

The text and workbook should be especially helpful as preparation for courses in experimental psychology and for courses in mental measurement in the field of education. Undergraduates who are taking—or who are planning to take—work in psychology need descriptive statistics, correlation and research techniques. Students in education (whether in teachers' colleges or elsewhere), whose interests lie primarily in the measurement of achievement and aptitudes, need

methods of treating test scores in addition to the standard statistical methods outlined above.

I am indebted to my colleagues, Professor Robert J. Williams and Dr. August A. Fink for a critical reading of the manuscript. And I am grateful to Miss Betty Jean Griswold for checking the problems in the text and in the workbook.

HENRY E. GARRETT

New York
January, 1956

CONTENTS

ELEMENTARY

STATISTICS

1.

STATISTICS AND MEASUREMENT

Why Study Statistical Method?

There are at least two reasons why students of the behavioral sciences need to study statistical method. The first is to enable them to read the literature, and the second, to perform class experiments or to carry out work on a research problem. The journals and the technical publications are full of statistical language. Even in the elementary textbooks the beginning student will encounter such statements as the following: the correlation between the intelligence test scores of offspring and mid-parent is about .50; 30% of sixth-grade boys exceed the median of sixth-grade girls in reading; the chi-square of 2.5, for one degree of freedom, yields a P which lies between .20 and .10, and hence is not significant; scores were normalized (expressed as T-scores) in order to make them equivalent; this survey employed stratified sampling, the sampling within the strata being random. Statements like these are well-nigh incomprehensible to the novice. Moreover, the student who skips the graphs, tables and formulas, in addition to skimming over the statistical description, must perforce rely upon the author's summary for such meager information as he is able to acquire.

Students in psychology and education cannot take advanced courses or carry out experiments without an elementary knowledge of statistics. Description in quantitative terms permits of a more precise summary. Moreover, statistical method, among other things, enables us to go beyond our result to a broader base, i.e., to generalize; to make predictions of probable achievement in school or vocation from test scores; to identify and evaluate factors that contribute to various aptitudes and personality traits; to discover the prevalence and strength of political and social attitudes and opinions.

Ways of Measuring

There are four levels at which mental and social measurement may be carried out. Beginning with simple *nominal* and *ordinal* arrangements, we move up progressively to the more precise *interval* and *ratio* scales. In nominal measurement, numbers are assigned to individuals or groups in order to distinguish them. Thus football players may be numbered 8, 25, 64, these designations serving to mark off one man from another; or sections of the same school grade are designated 1, 2, 3, etc. In ordinal measurement individuals or objects are put in 1-2-3 order for some quality or characteristic. Army officers may be arrayed in order-of-merit for demonstrated leadership; salesmen ranked on the basis of sales records or other criteria; children, for deportment; tonal combinations, for consonance or esthetic appeal. In ordinal arrangements there is no implication that the steps in the rank order are equal. Usually all we have is a serial arrangement running from high to low.

Interval-scales unlike nominal and ordinal arrangements have equal units or equal steps but no true zero point. Many mental tests are scaled in equal units (put into interval scales)

by one of several devices so that a 5-point gain from 40 to 45, say, is equivalent to a 5-point gain from 75 to 80. In mental measurement, however, a score of 40 is not *twice* a score of 20, as the reference point is not a true zero point of "just no ability." A young child, for example, may score zero on a test containing decimal fractions, not because he possesses zero ability in mathematics, but because the test is beyond his present educational level. If the test were extended down to include problems of a lower grade, he would doubtless achieve a score. The reference or zero point in interval-scales is usually an average, e.g., the mean or median (see p. 27).

Ratio-scales go a step beyond interval-scales: they have true zeros as well as equal units or steps. Measures of extent (in inches), of weight (in pounds), of time (in seconds), are illustrations of ratio-scales. A man six feet in height is three feet taller than a child three feet tall; moreover, the man is also *twice* as tall as the child, since measurement is from a true zero point. In physical measurement we make use of ratio-scales, but in mental measurements, except when expressed in time units, we must be content with interval-scales. In the behavioral sciences we deal mostly with ordinal arrangements and interval-scales.

The Meaning of Test Scores

Mental test scores may be expressed in two ways: as *amount done* in a given time, or as *time* taken to complete an assigned task. Time scores are used with tests requiring speed, in which the items are usually easy and all approximately equal in difficulty. Amount scores are the rule in power tests (in which the items increase in difficulty) as well as in inventories and questionnaires. The score is the number of correct answers or the number of items checked or marked in the time allowed. Mental test scores are conceived to be *distances*

rather than points along a behavorial yardstick. This is true although we usually express scores as integral numbers, e.g., as 12, 85, 226. Thus a score of 85 represents the interval from 84.5 up to 85.5, the score of 86 taking off from 85.5. The exact midpoint of score-interval 85 is shown in the diagram below:

Score of 85
84.5——85——85.5

If scores around 85 were graduated more finely, expressed as 84.8 and 85.3, for example, all such scores would fall on interval 85, and be recorded as 85, if expressed as two-place whole numbers.

STATISTICAL COMPUTATION

Rounding Numbers

The question of "how many places" to carry out a computation arises over and over again in statistical work. The number of decimals to be retained in an "answer" will always depend upon the nature of the problem: how accurate the data * were in the first place and hence how much accuracy is allowable in the result, whether a calculation is preliminary or final, and what it is to be used for. If we round off 12.83426 to two decimals it becomes 12.83; to one decimal, 12.8; to the nearest whole number, 13. A good general rule is to retain not more than two decimals in routine computation, as it is doubtful whether statistical work in the behavioral sciences often warrants greater accuracy. If the third decimal in a number is less than 5, drop it as shown above; if greater than 5, increase the preceding figure by 1 (e.g., 86.536 becomes 86.54); if equal to 5 exactly, compute a fourth decimal and

* The singular is datum. Data are figures, ratings, check lists and other information collected in experiments, surveys, and descriptive studies.

correct back to the second place (e.g., 86.5559 becomes 86.56); when exactly 5 followed by zeros, drop it and make no corrections (e.g., 92.35500 becomes 92.35).

Meaning of Significant Figures

If the height of a room is given as 12 feet, this result is said to be accurate to *two* significant figures; if recorded as 12.6 feet, accurate to three significant figures. We assume the measurement, 12.6 feet, to be correct to the nearest tenth of an inch, the true value lying between 12.55 and 12.65 feet. Two places to the *left* of the decimal point and one to the *right* are known: accordingly, 12.6 contains three significant figures, while 12 has only two significant figures. In general, the number of significant figures is an index of the accuracy of measurement and hence of the degree of confidence to be placed in our computation. The following examples should make clear the matter of significant figures.

386	has *three* significant figures.
386,000	also has *three* significant figures as it stands. The true value of this number lies between 385,500 and 386,500. Only the first three figures are fixed, the zeros serving simply to denote the size of the number. If the three zeros are known to be accurate the number has *six* significant figures.
3860.	has *four* significant figures. The decimal point fixes the size of the number and makes the zero significant.
.386	has *three* significant figures.
.3860	has *four* significant figures. The zero tells us that the fourth place is known to be zero.
.00386	has *three* significant figures. The first two zeros serve simply to locate the decimal.
9.000386	has *seven* significant figures. The integer 9 makes the three zeros significant: they are now measures and not merely markers.

Exact and Approximate Numbers

An exact number is one found by counting—15 boys, 20 books, 10 automobiles. An approximate number is a measure of some quantity; it is always subject to error, its accuracy depending upon the care with which the measurement is made and the precision of the measuring instrument. If a youngster is recorded as weighing 82 pounds, this figure probably means that his weight lies between 81.5 and 82.5 pounds. If his weight is recorded as 81.5 pounds, it probably lies between 81.25 and 81.75 pounds. In most cases it would be a waste of time to refine the measurement to this degree, however; in fact, 82 pounds is accurate enough for most purposes.

Test scores are always approximate numbers. Thus an *IQ* recorded as 126 implies a value between 125.5 and 126.5. Most scores are expressed as whole numbers, since greater accuracy is rarely warranted in test data. Computations based upon exact numbers offer no problem. They may be taken to as many decimals as one wishes, since the exact number 60 is really $60.000 \ldots n$. In calculations with exact and approximate values, the number of decimal places to be retained in the result is governed by the accuracy of the approximate number or numbers which enter into the calculation.

Computation Rules and Examples

(1) ACCURACY OF SUMS AND DIFFERENCES

Examples:

$8.6023 + 18.539 + 26.620 + 251.6 = 305.4$, rounded from 305.3613. The least accurate number (251.6) contains only one decimal: hence the result can have only one decimal.

263.91 − 150.626 = 113.28 rounded from 113.284. Here the less accurate number 263.91 contains two decimals. Hence the result can have two decimals.

Rule: The number of decimals to be retained in the sum or difference of approximate numbers should not be greater than the number of decimals in the least accurate number entering into the computation.

(2) ACCURACY OF A PRODUCT OR QUOTIENT

Examples:

$4.634 \times 152 = 704$, not 704.368, since 152 has only three significant figures. If 152 were written 152.0, we could write the result as 704.4 accurate to one decimal.

$5.87 \div .2685 = 21.9$, not 21.862, since 5.87 has only three significant figures.

Rule: The product or quotient of two approximate numbers can have no more significant figures than are present in the least (or less) accurate of the numbers entering into the computation.

(3) ACCURACY OF A SQUARE ROOT

Examples:

$\sqrt{824} = 28.7$, not 28.705, as there are only three significant figures in 824.

$\sqrt{44.6365} = 6.68106$, a result which would probably be rounded to 6.68. The six significant figures in 44.6365 permit us to have six significant figures in the root.

$\sqrt{24} = 4.8989795$ if 24 is an exact number. This root would almost certainly be rounded to 4.9.

Rule: The square root of an approximate number may legitimately contain as many significant figures as there are in the number itself. The square root of an exact number may be taken to as many decimals as one wishes.

Extracting Square Roots

Many students have difficulty with square roots even when a table of squares and square roots is available. The process is easy, however, if a few simple rules are followed. Suppose that we want the square root of 654,428. First, we must mark off the figures in pairs from *right* to *left* thus: 65/44/28. We know that the approximate square root must be 800, since $8^2 = 64$, and $800^2 = 64/00/00$. The nearest approximation to the square root of 654,428 that we can get from the table without interpolation is 809: this result is read from the column of numbers (first column) opposite 65/44/81 in the column of squares (second column). If 654,428 may be regarded as having six significant figures we are entitled to six significant figures in the root, and hence to three decimals. Should a more accurate root than 809 be desired, we must resort to interpolation in the table. The numbers in the square root table run only to 1000 and the simplest plan is to interpolate for the root between two squares in the squares column and read the root in the numbers column. The method is shown below:

Squares		*Numbers (Square root)*
65/28/64	1564	808
(65/44/28)	⟶	808.967
65/44/81		809
1617		

$$\frac{1564}{1617} = .967$$

The square root falls at 808.967, i.e., is .967 of the distance between 808.000 and 809.000, the roots lying just above and just below 65/44/28.

When numbers contain decimals, the figures are paired off

to the right and left of the decimal point. Thus the number 186.4321 is divided as follows: 1/86./43/21. The nearest integral root is the square root of 196 or 14. To find the closest approximation to the square root of 186.4321 without interpolation, we locate in the squares column the combination nearest to 1/86/43/21, disregarding decimals. This number is 1/87/69 and accordingly our approximation to the square root wanted is 13.7—not 137 or 1.37, as the closest integral root is 14. Our number, 1/86/43/21 is too large for the squares column in the table. Hence we must interpolate this time in the numbers column, if we want more places in our square root. The procedure is as follows:

Numbers		*Square root*
186		13.638
(186.4321)	$\longrightarrow$	(13.654, i.e., 13.638 + .016)
187		13.675
		.037

$.4321 \times .037 = .016$

The square root is .4321 of the distance (.037) between 13.638 and 13.675 or at 13.654. The table enables us to locate the square root to five significant figures, namely to 13.654. Rounded to two decimals, the square root is 13.65.

2.

THE FREQUENCY DISTRIBUTION

Organizing scores or other measures into the arrangement called a *frequency distribution* facilitates further statistical treatment as well as subsequent analysis and interpretation. Table 1(A) gives the scores achieved by 60 young men on the Army General Classification Test (AGCT)—a measure of general intelligence, administered to some 12,000,000 soldiers during World War II. In the lower half of the table the 60 scores have been grouped into 5-score *intervals* (i) or categories (sometimes called steps). This arrangement is a frequency distribution. Note that 3 scores (or a frequency of 3) fall in the top interval (120-124) which embraces the scores 120, 121, 122, 123, 124. The largest frequency, namely 15, is found in the middle interval (100-104); while only 3 scores fall in the bottom interval (80-84). From the frequency distribution as it stands, we know that most of our subjects are found in the middle of the scale (around score 100), relatively few achieving very high or very low scores.

Drawing Up a Frequency Distribution

The procedure to be followed in setting up a frequency distribution may be outlined as follows:

(1) First, determine the *range* or the distance from the highest to the lowest score. The highest score in Table 1 is 123 and the lowest is 81. Hence the range is 123-81 or 42.

(2) Next, settle upon the *number* and *size* of the intervals to be used in grouping the scores. Commonly chosen intervals are 3, 5, 10 units in length, as these are somewhat easier to work with in subsequent calculations. But *i*'s of 4, 7, and even 15 units are often encountered. A good working rule is to select a unit of classification (interval-length) which will yield from five to fifteen categories. This rule must sometimes be broken when the sample is very large or very small.

(3) Divide the range by the size of the *i* tentatively chosen. The number of *i*'s which a given range will yield can be found approximately (within one interval) by dividing the range by the *i*-sizes selected for trial. In Table 1(A) the range of 42

TABLE 1

Tabulation of 60 Army General Classification Test Scores into a Frequency Distribution with Interval = 5

A. Ungrouped Scores

97	107	96	118	81	113
*81	94	82	103	118	**123
107	93	98	102	97	106
86	92	93	112	104	115
121	99	103	100	110	107
104	103	100	108	102	104
89	100	111	85	97	100
104	117	109	104	122	98
112	99	90	100	91	92
96	105	95	114	109	87

* lowest score = 81
** highest score = 123
Range = 42

B. The same 60 AGCT scores grouped into a frequency distribution.

Intervals	*Tallies*	*f (frequency)*
120-124	|||	3
115-119	||||	4
110-114	~~||||~~ |	6
105-109	~~||||~~ |||	8
100-104	~~||||~~ ~~||||~~ ~~||||~~	15
95-99	~~||||~~ ~~||||~~	10
90-94	~~||||~~ ||	7
85-89	||||	4
80-84	|||	3
		$N = 60$

divided by 5 gives $8\frac{2}{5}$ and the number of *i*'s is actually 9. A unit of 2 yields 22 *i*'s * $(42/2 = 21)$; and a unit of 12 yields 4 *i*'s $(42/12 = 3\frac{1}{2})$. In the present example, a frequency distribution of 22 *i*'s would spread the data too thin, while a frequency distribution of 4 *i*'s would crowd the scores into too-large groupings. An interval of 5 was chosen, therefore, as being better suited to our data than an *i* of either 2 or 12. Furthermore, 9 *i*'s fall between 5 and 15 thus following the general rule in (2) above.

(4) List the intervals or steps as shown in Table 1(B). The top interval (120-124) actually begins at 119.5 (lower limit of score 120) and ends at 124.5 (upper limit of score 124) (see p. 6). More exactly, we could write this *i* as 119.5-124.5 and those below in the same way, thus:

Intervals	*f*
119.5-124.5	3
114.5-119.5	4
109.5-114.5	6
etc.	etc.

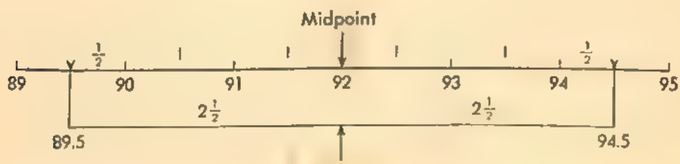
* One additional interval (122-123) will be needed to include the score 123.

Expressing i's in terms of their *exact* upper and lower limits is not generally as useful as is the method of writing interval-limits as *scores* (as shown in Table 1(B)). Writing score-limits is less time-consuming and avoids the confusion that arises when one i ends and the next begins with the same value, e.g., 114.5 (see above).

(5) Tally each score in its appropriate interval. The first score in Table 1(A), 97, falls in the i (95-99); the second score, 81, in the interval (80-84) and so on. When all 60 scores have been tabulated, the tallies on each interval are written as a single number in the third column under f (frequency) and the frequency distribution is complete. The sum of the f column—the sum of all the scores—is called N. In Table 1(B), N is 60.

The Midpoint of an Interval in the Frequency Distribution

When scores have been grouped into a frequency distribution they lose their separate identity and are represented by the midpoints of the i's upon which they fall. The midpoint of the interval (90-94) in Table 1(B), for example, is 92, as shown below:



The scale score of 92 lies 2½ score-units from the lower limit (89.5) and 2½ score-units from the upper limit (94.5) of the interval. An equation for computing the midpoint is as follows:

$$\text{Interval Midpoint} = \text{lower limit of } i + \frac{(\text{upper limit} - \text{lower limit})}{2}$$

In the example above, $\text{Midpoint} = 89.5 + \frac{(94.5 - 89.5)}{2} = 92.0$

When scores rather than exact limits are written as interval-limits (as is done in Table 1(B)) a simple rule for finding midpoints is

$$\text{Interval Midpoint} = \text{beginning interval score} + \frac{(\text{upper score} - \text{lower score})}{2}$$

or $\text{Midpoint} = 90 + \frac{(94 - 90)}{2}$ or 92.

Odd numbers (3, 5, 7) are usually to be preferred to even numbers (2, 4, 6) as interval-lengths, since they provide whole numbers as midpoints. To illustrate, consider the following i's wherein the units of classification are 4, 6, 7, 15, respectively: (20-23), (36-41), (67-73) and (135-149). The first i begins at 19.5 and ends at 23.5, and its midpoint is 21.5. The second interval has as its midpoint $36 + \frac{(41 - 36)}{2}$ or 38.5. The midpoint of the third interval is $67 + \frac{(73 - 67)}{2}$ or 70; and of the fourth is $135 + \frac{(149 - 135)}{2}$ or 142.

As a general rule, it is a good plan to begin the first interval with a multiple of the interval-size. If the lowest score is 26, for example, and the i selected is 5 units in length, begin with 25, making the first i (25-29). If the lowest score is 10 and the i selected is 3 units long, make the first interval (9-11).

Some Illustrative Frequency Distributions

The frequency distribution in Table 2 shows the number of errors made by a class of 20 students in five consecutive trials

on a pencil maze. Contrary to the order of the *i*'s in Table 1(B), the *best* score (i.e., 0) is the lowest numerically and the *poorest* score (i.e., 6) is the highest, since scores are expressed in terms of errors made. The range is 6 (6-0) and the *i* size is 1.

TABLE 2

Errors Made by a Class of 20 Students in 5 Consecutive Tracings of a Pencil Maze

Scores (errors)	*f*
6	2
5	4
4	4
3	3
2	4
1	2
0	1
	$N = 20$

The true limits of the first interval are 5.5 to 6.5, 6 being the midpoint. The exact limits of the first and of the other intervals in the table may be written as follows:

Scores (errors)	*f*
5.5-6.5	2
4.5-5.5	4
3.5-4.5	4
2.5-3.5	3
1.5-2.5	4
.5-1.5	2
–.5- .5	1
	$N = 20$

Note that the bottom interval extends from —.5 to .5 with 0 as the midpoint. It is apparent that the error scores shown in Table 2 are actually the midpoints of unit-intervals.

Table 3 shows two additional frequency distributions. The first (A) represents the distribution of *IQ*'s for 660 runaway boys. The interval size is 10 and the number of *i*'s is 9. Note that the midpoints of the first three *i*'s are 114.5, 104.5 and 94.5. The *i*-size is an even number (10) and hence the midpoints are fractions. In each case, 5 (½ of the *i*) may be added to the actual lower limit of the interval to give the midpoint; or more easily 4½ may be added to the lower *score* limit. Thus the midpoint of interval (100-109) is 100 + 4½ or 104.5. It may be noted in passing that 208 of the 660 boys have *IQ*'s which fall on intervals *below* (70-79). These *i*'s are (60-69) to (30-39) inclusive. An *IQ* below 70 is usually taken to be indicative of feeblemindedness. This means, therefore, that nearly ⅓ of these runaway boys must be classified as feebleminded in terms of this criterion.

TABLE 3

A. Frequency distribution of the *IQ*'s of 660 runaway boys [Armstrong, C., *660 Runaway Boys,* R. C. Badger (Boston: Gorham Press, 1932), p. 31]. $i = 10$

IQ intervals	*Midpoints*	*f*
110-119	114.5	14
100-109	104.5	37
90-99	94.5	78
80-89	84.5	139
70-79	74.5	184
60-69	64.5	140
50-59	54.5	60
40-49	44.5	6
30-39	34.5	2
		$N = 660$

B. Frequency distribution showing the numbers of pupils in 43 classrooms of three elementary schools. The f is the number of rooms containing the class sizes shown in the first column. $i = 3$

Number of pupils	*Midpoints*	f
36-38	37	2
33-35	34	5
30-32	31	17
27-29	28	10
24-26	25	8
21-23	22	1
		$N = 43$

In the second frequency distribution in Table 3(B), each interval covers 3 units. This frequency distribution shows that two rooms in these three elementary schools have 36-38 pupils, five rooms 33-35 pupils and only one room 21-23 pupils. Reading down the first column, the midpoints are 37, 34, 31, 28, 25, 22. The odd-sized interval causes the midpoints to be whole numbers.

GRAPHICAL METHODS

It has often been said that "one picture is worth 1000 words." And to a lesser degree, perhaps, but for the same reasons, a diagram or graph owing to its greater vividness and comprehensiveness is often more revealing than the most careful array of numbers. There are two ways in which a frequency distribution may be represented graphically: (1) by a *frequency polygon* * and (2) by a *histogram*. These two graphs together with another much-used device, the *line graph*, will be described in this section.

* Polygon means many-sided figure.

The Frequency Polygon

Figure 1 shows a frequency polygon of the 60 AGCT scores tabulated in Table 1(B). First, two axes, a horizontal or X-axis and a vertical or Y-axis have been drawn at right angles. Along the X-axis or *base line*, score intervals are then laid off

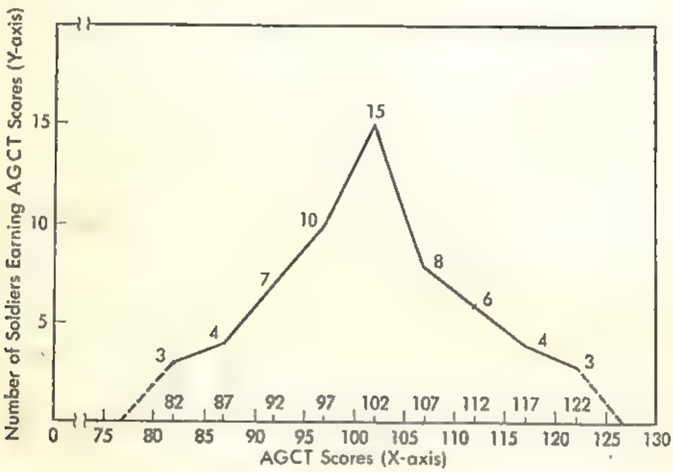


Figure 1

at regular distances from 80, the lower score limit of the first interval. The break in the X-axis ($\int\int$) indicates that the vertical or Y-axis has been moved in for convenience. The 3 scores on the bottom i (80-84) are represented by a point just above the midpoint of the interval, namely, 82, and 3 units up on the Y-axis. The 4 scores on the next interval (85-89) are represented by a point 4 units above 87, the midpoint of this i, and so on for the others. Thus the largest f, namely 15, lies just above 102, the midpoint of i (100-104).

When all of the points have been located and marked off, they are joined by a series of short lines to give the outline of

the frequency polygon. In order to complete the figure, i.e., to bring it down to the base line or X-axis, two i's have been added, one at the low end (75-79) and one at the high end (125-129) of the distribution. The frequency on each of these i's is, of course, zero; and accordingly, the midpoints lie on the X-axis. The addition of these two extreme i's enables us to have the frequency distribution begin and end on the X-axis.

In order to provide a symmetrical figure—one which is neither too squat nor too thin—units must be selected carefully for the two axes. A good rule is to select units which will make the height of the frequency polygon roughly ⅔ to ¾ of its width. In Figure 1 a unit was selected for the score intervals on the X-axis which would fit comfortably on the page. A Y-unit was then chosen which would make the height of the figure at 15 (the maximum frequency) about ⅔ of the width of the frequency polygon. This can be done very simply in the following way. There are 10 intervals, counting the 2 half-intervals at the extremes on the X-axis. To follow the ⅔ rule, the peak of the figure (maximum height) should, therefore, be equal roughly to 6-7 X-axis intervals. This is the height represented by a frequency of 15. Intermediate points on Y can readily be fitted in from 0 to 15, slight adjustments being made to fit the graph paper. The total frequency (namely, N) of the distribution is represented by the *area* of the frequency polygon: the area bounded by the broken lines of the frequency surface and the X-axis.

The Histogram or Column Diagram

The frequency distribution of Table 1 is shown again in Figure 2 in the form of a histogram, sometimes called a column diagram. The histogram differs from the frequency polygon in several ways. In the first place, instead of a single dot over the midpoint of an interval (i) to represent the fre-

quency thereon, a small rectangle is drawn, its height equal to the frequency on the *i*, and its width equal to the width of the *i*. There is no need to add *i*'s at the extremes of the frequency distribution in order to close the figure. In Figure 2, the first *i* begins at 79.5, the actual lower limit of the interval (80-84) and ends at 84.5, the upper limit of the *i*. The 3 scores on this interval are represented by a rectangle stretching from 79.5 to 84.5 and 3 units high. Note again that there is a break ($\int\int$) to the left of 79.5 to indicate a shortening of the base line.

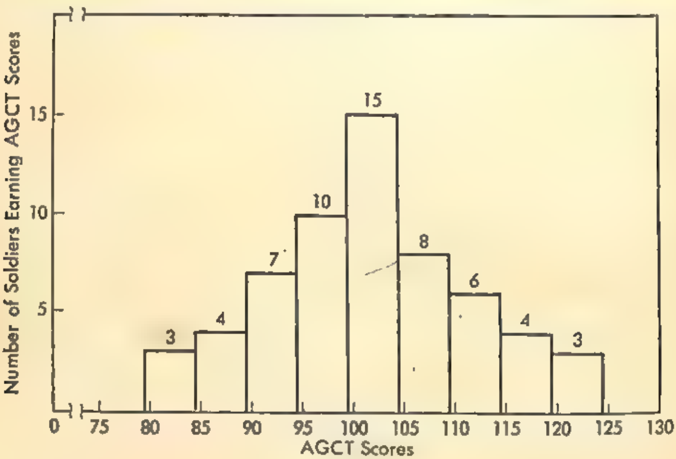


Figure 2

The total frequency (N) of the distribution is again represented by the area of the figure. In the histogram, moreover, the area of the small rectangle placed over each interval is directly proportional to the f on that interval. This relationship of proportionality is not true of the frequency polygon owing to boundary irregularities from point to point. The height of the histogram should be roughly ⅔-¾ of its width in order to provide a symmetrical figure.

Figure 3 pictures a histogram of the IQ distribution of the 660 runaway boys taken from Table 3. The figure is quite

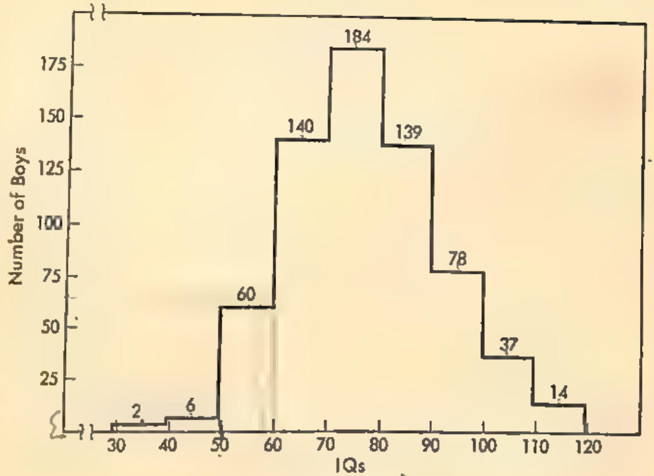


Figure 3

symmetrical. Note that the rectangle over the first *i* begins at 29.5 and ends at 39.5. Points along the base line could be marked off at 29.5, 39.5, 49.5, etc., but it is perhaps easier to mark off points at 30, 40, 50, etc. In Figure 4 the distribution

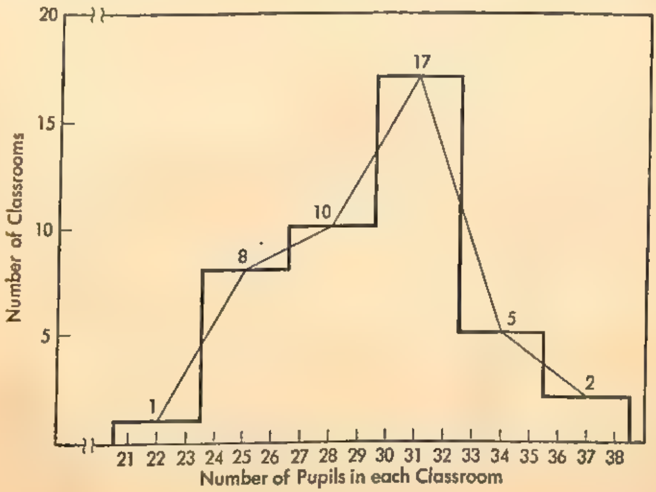


Figure 4

showing the frequency of classroom sizes in Table 3(B) is represented both as a histogram and as a frequency polygon on the same axes.

Comparison of the Frequency Polygon and Histogram

Both these graphical devices tell the same story and there is little to choose between them. The areas of rectangles over the i's in the histogram are directly proportional to the f on the i; and the figure does not have to be closed, as does the frequency polygon, by adding intervals at the two extremes of the distribution. This is sometimes an advantage. The frequency polygon is to be preferred to the histogram when two groups—for example, sixth grade boys and sixth grade girls—are compared in terms of test performance on the same X- and Y-axes. In the histograms the vertical and horizontal lines often coincide and are difficult to disentangle. This will happen less often with frequency polygons.

Line Graphs

The line graph is especially useful when an experimenter wants to show the trend of performance or other happenings over a period of time, over successive trials, for different age levels, and so on. Figure 5 represents the number of errors made by one person in typing by the touch system over nine successive days. Days (trials) are marked off at regular intervals along the base line or X-axis and scores showing the number of errors are indicated by points in inverse order on the Y-axis above the successive days. Since the errors *decrease* with practice, the rising curve shows a steady *increase* in accuracy. Units selected for the horizontal and vertical axes should not stretch out the line graph too much nor make it

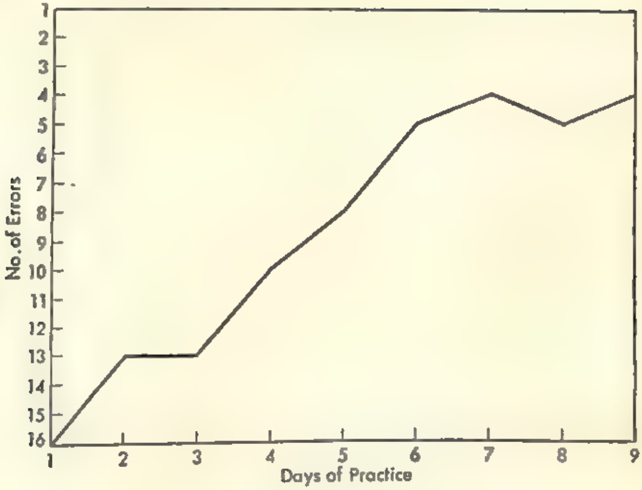


Figure 5

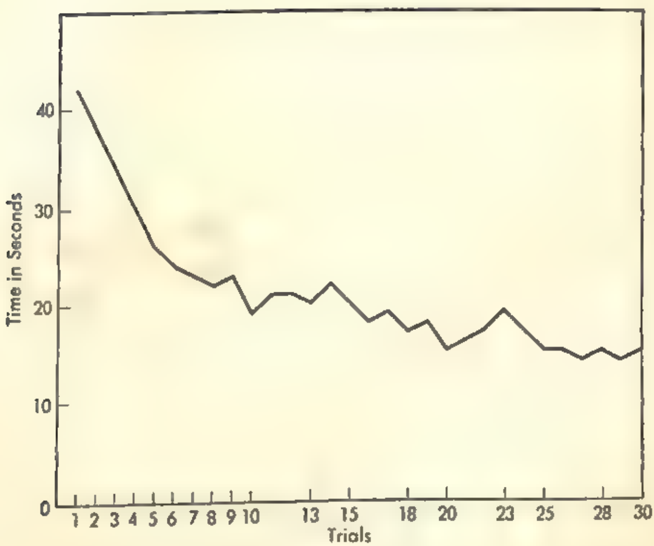


Figure 6

appear crowded. The ⅔-¾ rule will be helpful here; have the maximum height of the graph about ⅔-¾ of its width.

Figure 6 represents the progress made in learning to write the alphabet backward. Successive trials are marked off on the *X*-axis, and performance in terms of seconds taken to complete each trial on the *Y*-axis. The time needed to perform the task drops sharply from the first to the fifteenth trial; and more slowly thereafter up to the thirtieth trial.

Each axis of a line graph should be carefully labeled, e.g., as trials, times, errors, scores and the like. Successive values of *X* and *Y* should be marked off clearly on the axes so that the reader will know the size and number of the units employed.

3.

AVERAGES

An average or "measure of central tendency" is a value which is typical of a number of like things. We speak of the average grade of a class on a history test, the average height of ten-year-old boys, the average price of a stock over the last five years, meaning in each instance a value representative of the entire set of things being considered. There are three sorts of averages in common use, the *arithmetic mean*, the *median* and the *mode*. The arithmetic mean or more simply the mean (M) is popularly spoken of as "the average." The other two averages are less generally encountered outside of technical reports making use of statistical method.

THE MEAN

The Mean Calculated from Ungrouped Scores

If a workman earns \$8, \$6, \$10, \$9, and \$12 per day over a five-day week, his "average" wage per day is \$9 (\$45/5 = \$9). This average or mean is defined simply as the sum of a set of measures or scores divided by their number. The formula for the mean is

$$M = \frac{\Sigma X}{N} \tag{1}$$

(the arithmetic mean from ungrouped scores)

where

M = the arithmetic mean
X = a score or other measure
N = the number of scores
and Σ (the Greek letter "sigma") denotes "sum of."

In Table 4 below the mean by formula (1) is 6.4.

TABLE 4

Calculation of M from 10 Ungrouped Measures. The data are the scores achieved on a memory span test by ten seven-year-olds

Scores (X): 5, 6, 8, 5, 7, 8, 6, 9, 5, 5

$\Sigma X = 64$
$N = 10$
$M = 64/10 = 6.4$

The Mean Calculated from a Frequency Distribution

When scores or other measures have been put into a frequency distribution the M may be computed in either of two ways: (a) *directly* by using the midpoints of the successive intervals as scores, or (b) *indirectly* by first assuming a mean and treating the i as a unit. These two procedures will be described separately.

(1) CALCULATION OF THE MEAN USING THE MIDPOINTS AS SCORES

When scores have been grouped into a frequency distribution the individual measures lose their identity and are repre-

sented by the midpoints of the i's on which they fall. To find the M, therefore, we must multiply the midpoint of each i by the f upon it; add the fX (the X are now midpoints) to get ΣfX and divide by N. The formula is

$$M = \frac{\Sigma fX}{N} \tag{2}$$

(the mean when scores are grouped into a frequency distribution)

where

M = the arithmetic mean
ΣfX = sum of the frequencies times the midpoints of their respective i's
N = the number of cases

The two distributions in Table 5 (the data are from Table 3, p. 18) will serve to illustrate formula (2) and the method of direct calculation. In the first frequency distribution the 660 IQ's have been grouped into 10-unit i's. The midpoints are listed in the second column. To find the mean, multiply each midpoint (X) by the frequency on that i and divide the sum [ΣfX] by N. Thus ($50930 \div 660$) gives a mean of 77.17. As the interval lengths are even (10 units) each midpoint contains the decimal .5.

In the second example, the data represent the number of pupils enrolled in each of 43 classrooms. Again the midpoints of the i's are multiplied by the appropriate f on each i to give the fX. The ΣfX divided by N ($1273 \div 43$) gives an M of 29.60. This is the "average" class size. As the i-size is odd (3 units), the midpoints this time are integers.

(2) CALCULATION OF THE MEAN BY THE ASSUMED MEAN METHOD

Although not immediately apparent to the beginning stu-

TABLE 5

Calculation of the Mean from a Frequency Distribution

A. Data reporting the *IQ*'s of 660 runaway boys (see Table 3)

IQ intervals	*Midpoints (X)*	*f*	*fX*
110-119	114.5	14	1603.0
100-109	104.5	37	3866.5
90-99	94.5	78	7371.0
80-89	84.5	139	11745.5
70-79	74.5	184	13708.0
60-69	64.5	140	9030.0
50-59	54.5	60	3270.0
40-49	44.5	6	267.0
30-39	34.5	2	69.0
		$N = 660$	50930.0

$$M = \frac{\Sigma fX}{N} \text{ by formula (2)}$$

$$= \frac{50930}{660}$$

$$= 77.17$$

B. Data showing the number of pupils in 43 classrooms in three elementary schools (see Table 3)

Number of pupils (intervals)	*Midpoints (X)*	*f*	*fX*
36-38	37	2	74
33-35	34	5	170
30-32	31	17	527
27-29	28	10	280
24-26	25	8	200
21-23	22	1	22
		$N = 43$	1273

$$M = \frac{\Sigma fX}{N} = \frac{1273}{43} = 29.60$$

dent, the calculation of the *M* by the assumed mean method has several advantages over the direct method just outlined. In the first place, the assumed mean method requires less computation and is considerably less time-consuming when *N* is large. Again, the assumed mean method provides a computation model which will be followed when the standard deviation and the correlation coefficient are calculated later on. The sooner the student learns the method, therefore, the better.

Table 6 provides two illustrations of the assumed mean method. References will be made to these examples in the following outline of computation.

(a) First assume or "guess" a mean (*AM*) preferably on the *i* having the largest *f* and as near to the middle of the distribution as possible. This procedure reduces the computation, but it should be noted that the method works no matter upon what *i* the *AM* is taken. In Table 6(A), the *AM* is taken at 74.5, midpoint of *i* (70-79), and there are 184 *f*'s on this *i*. In Table 6(B) the *AM* is taken at 31, midpoint of interval 30-32. This *i* also has the largest *f*, namely, 17.

(b) In the column headed x' list the deviations of each *i* midpoint from the *AM* in units of interval. In the first example, midpoint 84.5 deviates 10 scores or 1 interval from 74.5 (*AM*); 94.5 deviates 2 intervals; 64.5 deviates -1 interval and 54.5 deviates -2 intervals. In each instance, $x' = X - AM$, where *X* is the midpoint and *AM* the assumed mean. The prime ($'$) shows that the deviation is from the *AM*. When the deviation is from the *M* and not the *AM* it is written x and not x'.* In example (B)—the midpoints need not be written in—there are 2 interval deviations above and 3 below the *AM*.

* It can be shown mathematically that the sum of the deviations around the mean is zero: i.e., $\Sigma(X - M) = \Sigma x = 0$. The size of the $\Sigma fx'$ tells us the amount by which the *AM* misses the *M*—plus or minus. The net fx', therefore, provides a correction (c) to be applied to the *AM*.

(c) Multiply each f by its corresponding x' to give fx'. [See Table 6(A).] Now add the plus and minus fx' separately and find the absolute difference, attaching the sign of the larger sum. From the $\Sigma fx'$ we find the correction (c) in interval-units which must be applied to the AM to give the M. The formula for c is

$$c = \frac{\Sigma fx'}{N} \text{ (algebraic)}$$

As shown in the table, the correction for example (A) is 176/660 or .2667. When this c is multiplied by i (the length of the interval), we have ci, the correction in score-units. Thus .2667 × 10 gives 2.667; and this correction added to

TABLE 6

Calculation of M by Assumed Mean Method

A. Data from Table 3: 660 *IQ's*

IQ intervals	*Midpoints*	f	x'	fx'	
110-119	114.5	14	4	56	
100-109	104.5	37	3	111	
90-99	94.5	78	2	156	
80-89	84.5	139	1	139	462
70-79	74.5	184	0		
60-69	64.5	140	−1	−140	
50-59	54.5	60	−2	−120	
40-49	44.5	6	−3	−18	
30-39	34.5	2	−4	−8	−286
		$N = 660$			Diff. = 176

$$AM = 74.5$$

$$c = \frac{\Sigma fx'}{N} = \frac{176}{660} = .2667$$

$$ci = .2667 \times 10 = 2.667$$

$$\begin{aligned} M &= AM + ci \\ &= 74.5 + 2.667 \\ &= 77.17 \end{aligned}$$

B. Data from Table 3: 43 classrooms

Number of pupils (intervals)	f	x'	fx'	
36-38	2	2	4	
33-35	5	1	5	9
30-32 (31)	17	0		
27-29	10	−1	−10	
24-26	8	−2	−16	
21-23	1	−3	−3	−29
	$N = 43$			Diff. = −20

$$AM = 31$$

$$c = \frac{\Sigma fx'}{N} = \frac{-20}{43} = -.4651$$

$$ci = -.4651 \times 3 = -1.3953$$

$$M = AM + ci$$
$$= 31 - 1.40$$
$$= 29.60$$

74.5, the *AM*, equals 77.17, the mean. The formula for the *M* is

$$M = AM + ci$$

When the sign of the *ci* is minus, it is subtracted from the *AM;* when plus, added.

In the second example, the fx' is −20 (−29 and 9) and the correction (c) is −20/43 or −.4651. When this correction (in units of i) is multiplied by 3 (the interval-length) we have $ci = -1.3953$ ($-.4651 \times 3$). The *M* is now 31 − 1.40 (to two decimals) or 29.60. Note that the two *M*'s found by the *AM* method check exactly the *M*'s found by the direct method (Table 5).

The Midpoint as Representative of the Scores on the Interval

In both of the procedures outlined above for computing the mean from a frequency distribution, the midpoint is taken to

be representative of *all* of the scores on a given interval. When the i's are large (10-15, for instance) and N is small, the f on an interval may not be distributed symmetrically about its midpoint. In i's above the mean, for example, the frequencies tend to lie *below* the midpoints more often than above; while in i's below the M, the f tend to lie *above* the interval midpoints more often than below. These two tendencies will, in general, cancel each other out when an M is calculated from *all* of the intervals. Hence the so-called grouping error can usually be safely ignored when an M is calculated from a frequency distribution.

The Mean as a Point

The mean is always a point along some continuum or yardstick, and is not a score, so that the term mean is the correct one, not "mean score." Usually the M is a mixed decimal, for example, 28.64 or 324.81; when it comes out as a whole number (12 or 126) it is written 12.00 and 126.00. The fact that the M is a point on a scale often leads to results which seem unreal to the beginner. We read, for example, that over a given period the mean number of children in the families of college graduates is 1.73; or that there is .85 of an auto per family in a certain city. There are, of course, no fractional children nor fractional automobiles. But the result is reasonable when the M is thought of as a *point* along a scale—a point which expresses the center of density within the distribution as a whole.

THE MEDIAN

Like the *mean,* the *median* is also an "average." It is defined simply as that *point* in the distribution of scores above which (or below which) lies 50% of the frequency (N). While the

median is also a point, unlike the mean it is found by counting into the distribution, not by summing the scores (X) and dividing by N. Sometimes the median is wanted when the scores are ungrouped, but usually it is calculated from a frequency distribution. We shall, therefore, first consider the calculation of the median (Mdn) from a frequency distribution and return later to the question of what to do with ungrouped scores.

Calculation of the Median from a Frequency Distribution

Table 7 illustrates the computation of the Mdn from the frequency distributions shown in Tables 1 and 3. The steps in computation may be set out as follows:

(1) First cumulate the f's from bottom to top of the distribution. This is not necessary, but it aids computation and prevents errors.

TABLE 7

Computation of Mdn from frequency distribution

A. Data from Table 1: 60 AGCT scores

(1) *Intervals*	(2) f	(3) *cum f*	
120-124	3	60	
115-119	4	57	
110-114	6	53	
105-109	8	47	
100-104	15	39	24
95-99	10	24	
90-94	7	14	
85-89	4	7	
80-84	3	3	
	$N = 60$		
	$\frac{N}{2} = 30$		

$$Mdn = 99.50 + \frac{6}{15} \times 5$$
$$= 99.50 + 2$$
$$= 101.50$$

B. Data from Table 3: 660 *IQ*'s

(1) *IQ intervals*	(2) *f*	(3) *cum f*	
110-119	14	660	
100-109	37	646	
90-99	78	609	
80-89	139	531	
70-79	184	392	208
60-69	140	208	
50-59	60	68	
40-49	6	8	
30-39	2	2	
	$N = 660$		

$$\frac{N}{2} = 330$$

$$\begin{aligned} Mdn &= 69.50 + 10\left(\frac{330 - 208}{184}\right) \text{ by formula (3)} \\ &= 69.50 + 6.63 \\ &= 76.13 \end{aligned}$$

(2) Take ½ of N and count into the cumulative distribution (col. 3) from the low end until the value next greater than ½ N is reached. In example (A), ½ N is 30, and since 39 is the first cumulative f larger than 30, we must stop at 24. The 24 scores counted off take us through interval 95-99 and up to 99.5, the actual beginning of interval 100-104, the interval which must contain the *Mdn*. In order to count off the additional 6 scores necessary to bring 24 up to 30 and thus reach the *Mdn*, we take $6/15 \times 5$ (interval) and add the result (namely, 2) to 99.5. This locates the *Mdn* at 101.50. Note that the 6 additional scores are divided by 15 (f on interval 100-104); and this fraction of 5 (the length of the i) tells how far we must go into (100-104) to reach the *Mdn*.

(3) A formula for the median which includes all of these calculations is

$$Mdn = l + i\left(\frac{\frac{N}{2} - cum\, f_l}{f_m}\right) \tag{3}$$

(the *Mdn* calculated from a frequency distribution)

where

l = lower limit of i upon which the *Mdn* lies
$N/2$ = ½ of the total number of scores (N)
$cum\, f_l$ = sum of scores on i's below l
f_m = frequency on the i containing the *Mdn*
i = length of interval

Applying the formula to example (A) in Table 7, we have

$$Mdn = 99.5 + 5\left(\frac{30 - 24}{15}\right)$$
$$= 101.50$$

When formula (3) is applied to the data of example (B) in Table 7, the

$$Mdn = 69.50 + 10\left(\frac{330 - 208}{184}\right)$$
$$= 69.50 + 6.63 = 76.13$$

We are able to count off 208 in the *cum f* column without exceeding $330\left(\frac{N}{2}\right)$. This takes us up to 69.5, lower limit of interval (70-79) which contains the median. In order to count off the 122 (330-208) additional scores needed to reach 330, we must take $\frac{122}{184} \times 10$ (interval) and add the result (6.63) to 69.5 to locate the median at 76.13. Formula (3) is useful and not difficult to apply, but sometimes it is easier to count $\left(\frac{N}{2}\right)$

directly into the distribution as was done in the example (A) above.

Calculation of the Median from Ungrouped Scores

When scores are few in number and are ungrouped, they may be arranged in order of size and the *Mdn* found by counting off ½ of *N* from either end of the series. This procedure is simple in principle, but it presents difficulties when several scores are repeated or when there are gaps in the order—scores missing. It is usually easier and more accurate, therefore, to put the data into a frequency distribution and compute the median by the method used above; and this procedure is generally to be recommended. To illustrate, suppose we want the *Mdn* of the 10 memory span scores shown in Table 4, page 28. These scores may be grouped into intervals of 1 (5 is 4.5-5.5, 6 is 5.5-6.5 and so on) as follows:

TABLE 8

Intervals	*f*	*cum f*
9	1	10
8	2	9
7	1	7
6	2	6
5	4	4
	$N = 10$	
	$\frac{1}{2} N = 5$	

By formula (3) the *Mdn* is

$$Mdn = 5.5 + 1\left(\frac{5-4}{2}\right)$$
$$= 6.0$$

From the frequency distribution, the *Mdn* is readily found to be 6.0. If the 10 scores are listed in order of size we have:

5 5 5 5 6 6 7 8 8 9. Counting off 5 scores from the beginning of the list, we count through 6—or up to 6.5, upper limit of the score 6. Counting off 5 scores from the upper end of the series, we again count through 6, but this time down to 5.5, the lower limit of the score 6. The point midway between 5.5 and 6.5 is 6.0 and this point is recorded as the *Mdn*. Grouping the data into a frequency distribution is less ambiguous and offers less chance for error.

THE MODE

Like the *mean* and the *median*, the *mode* is also an "average." It is defined simply as the most common—oft-recurring—measure or score in a series; that is, it is the "most popular" measure. In a frequency distribution the mode is usually taken as the midpoint of that *i* which contains the largest frequency. This midpoint mode is often called the "crude mode" to distinguish it from the "true" or theoretical mode, the point which describes the actual peak in the distribution. In Table 7 the crude mode in example (A) is 102.0, midpoint of (100-104); in example (B) the crude mode is 74.5.

A formula for approximating the true mode is

$$Mode = 3\ Mdn - 2\ Mean \tag{4}$$

(approximation to the true mode in a frequency distribution)

For example (A) in Table 7, the mode by formula (4) is

$$\begin{aligned} Mode &= 3 \times 101.50 - 2 \times 101.67 \\ &= 101.16 \text{ as against a crude mode of } 102.0 \end{aligned}$$

The mode for example (B) is

$$\begin{aligned} Mode &= 3 \times 76.13 - 2 \times 77.17 \\ &= 74.05 \text{ as against a crude mode of } 74.5 \end{aligned}$$

The mode is often employed as a simple "inspectional average"—to provide a rough notion of the concentration of scores. For this purpose the crude mode is usually sufficient.

When to Use the Mean, the Median, and the Mode

As we have said above, the mode is most often used as a preliminary measure of central tendency, and hence it is rarely necessary to choose between it and either the *M* or *Mdn*. It is often a real question, however, whether the mean or median is to be preferred. The general rules given below will help in making a decision in many specific cases.

(1) Use the mean when the most *stable* measure of central tendency is wanted. The *M* has a smaller standard error (see p. 91) than the median and is less variable from sample to sample.

(2) Use the mean when the size of each score should enter in and influence the central tendency [see (4) below].

(3) Use the *M* when *SD*'s and correlation coefficients—statistics computed from the *M*—are to be found later.

(4) Use the *Mdn* when there are extreme scores at either end of the series. In the simple array of 5 scores—12, 18, 25, 25, 40—the mean is 24 and the median is 25. If 1 extra large score, say 120, is added, however, the *M* becomes 40 while the *Mdn* is still 25. In this case, the *Mdn* is a better indicator of the typical score, the mean being unduly influenced by the single large score. To take another illustration of the same sort, suppose that in a church collection ten people give 5 cents, twenty give 10 cents, and that one affluent sinner contributes $50.00. The mean is $1.69—a completely unrealistic measure of the typical donation. The *Mdn*, however, is 10 cents, which is certainly much closer to the "average" amount donated.

(5) Use the *Mdn* when certain scores should influence the

measure of central tendency, but all that we know about them is that they lie outside the distribution. Children sometimes fail to finish a test in the allotted time and are marked *DNC* (did not complete); or a very large score may "run off the scale" at the upper end of the distribution. *DNC*'s, large scores and zero scores all have a weight of one (the same as other scores) when we count in to find the median, and hence do not affect this measure unduly. But *DNC*'s cannot be added in to give the mean, and when scores are very large or very small (e.g., zero) they affect the mean markedly, as we saw in (4) above.

There are occasions when neither the mean nor the median is an adequate measure of central tendency. The distribution, for example, may be *bimodal*—show two distinct peaks, one at the high and one at the low end of the scale. Or the distribution may even be *multimodal*—have more than two peaks. In such cases, the median or the mean or both may fall into a valley between two peaks (see Workbook, p. 9, for a two-peaked distribution); neither gives a fair picture of score concentration. When the frequency distribution has two or more peaks, we should first check for computational errors. Next, we might get a larger sample—increase N. If neither of these procedures removes the multimodality, we should simply represent the distribution graphically and give no measure of central tendency.

In summary, final decision as to what measure of central tendency should be employed must depend (1) upon the character of the data—whether ratings, scores, proportions, etc.—and the form of the distribution; (2) upon the accuracy desired—the mode may give a sufficient indication; and (3) upon the purpose for which the measure of central tendency is calculated. In experimental and research studies, a higher level of accuracy and greater precision are required than in routine statistical descriptions.

When distributions are symmetrical around a central peak (see Figure 1, p. 20) the M and Mdn fall at the same or almost the same point on the scale. Off-center or non-symmetrical distributions are said to be *skewed;* and the skewness may be either to the *right* or to the *left.* The discrepancy between the M and Mdn is often taken as a measure of the amount and the direction of skewness (see p. 85).

4.

VARIABILITY

We have all observed that within any group—students, factory workers, clerks, even ministers—there are large individual differences from one person to another. The members of a given group differ in such objective characteristics as height, age, physical strength and appearance. And they differ also in less readily perceived—but perhaps more important—attributes, such as industry and intelligence as well as in social and personality traits.

Knowing the *variability* of performance within a group may be more valuable than knowing the typical performance (average) of the group. After we have computed a measure of central tendency (mean or median), therefore, usually the next step is to calculate a measure which will show how much the group spreads or scatters around this typical point.

An illustration will show the value of a measure of variability. Suppose that two sections in college algebra have been set up in the same school. Let us assume that aptitude for algebra is equal in the two groups—as shown by mean score on a general mathematics test given at the start of the term—but that the scatter of scores is widely different. Figure 7 represents the situation graphically. The *M*'s of the two groups are equal—both at 60—but in group *A* the range of individual scores is from 20 to 100, whereas in group *B* the range of scores

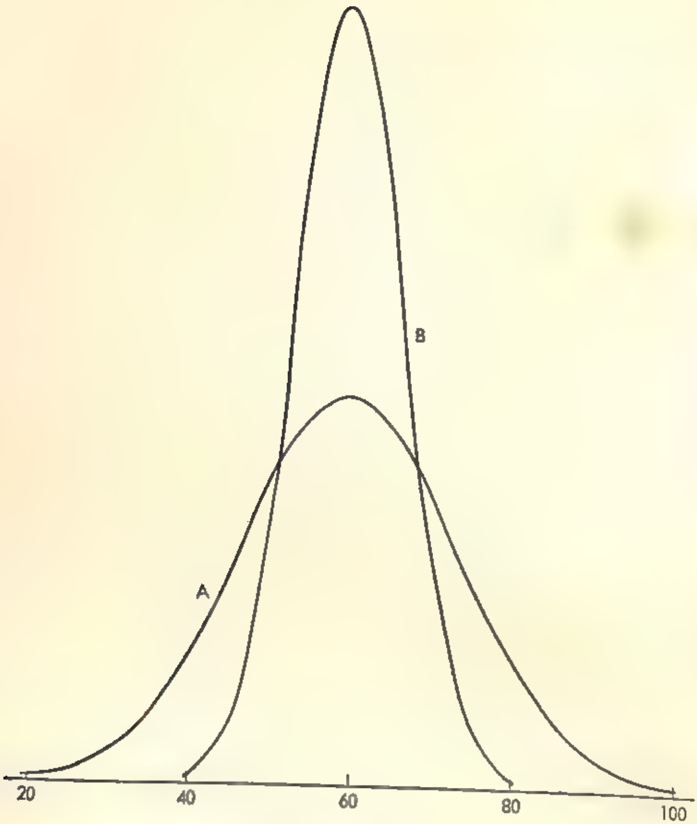


Figure 7

is from 40 to 80. The *A* group covers twice as much distance on the score-scale as the *B* group and contains very able as well as very inept students. In contrast, the students in group *B* are relatively homogeneous in mathematical ability.* A large group of factory workers will contain individuals who are from three to four times as competent as the poorest worker in terms of output, speed, dexterity and so on. Even

* It would be much easier to teach the *B* than the *A* group. If instruction were geared to the so-called average student (of score 60) most of the *B*'s might profit. But the top students in *A* would be bored and the bottom students confused—and perhaps dismayed.

within a highly selected group of artists, musicians, or graduate students wide variations in ability will be common.

Variations in performance are found, of course, within the *same* individual as well as within a group. In fifty finger reactions to a visual stimulus, for example, a student's mean reaction time may be 120 ms. whereas his individual reactions range from 100 to 140 ms.* Variations within the same person are often called *intra*-individual differences to distinguish them from *inter*-individual differences—differences from person to person within the same group. The frequency distribution of the measures taken from a single person is treated in the same fashion as is the frequency distribution for a group all of whom have taken the same test.

Four statistical measures have been devised to represent the variability or dispersion within a set of scores or other measures of behavior. These are (1) the range, (2) the quartile deviation, (3) the standard deviation, and (4) the average or mean deviation. We shall take up the calculation of the first three of these in some detail, postponing treatment of the mean deviation until page 61.†

THE RANGE

We have already had occasion to use the range (p. 13) in determining the number and size of the intervals to be used in a frequency distribution. The range may be defined as the difference between the largest and smallest scores. In Table 1, the range is 42 (123-81); in Table 2 it is 6 (6-0). When scores have been organized into a frequency distribution an approximate range is given by subtracting the lower limit of the bottom interval from the upper limit of the top interval. In

* ms. means milliseconds (thousandths of a second).

† The average or mean deviation (called *AD* or *MD*) is so rarely used as to be virtually obsolescent.

Table 3(A), for example, the range is approximately 90 (119.5 — 29.5); and in Table 3(B), the range is approximately 18 (38.5-20.5). When scores are widely scattered and especially when there are gaps at the extremes of the distribution, the range is likely to be an inefficient measure of variability. Suppose, for example, that the highest score in a distribution is 82 and there is a gap of 12 points before we reach 70, the next score. Now if the lowest score is 40, the single high score of 82 will increase the range from 30 (70-40) to 42 (82-40).

The range is most often used as a preliminary measure of spread in the scores, and for this purpose its lack of precision is not especially serious.

THE QUARTILE DEVIATION OR Q

The quartile deviation—called Q—is defined simply as one-half the distance between the 75th and the 25th percentiles in a frequency distribution. The 25th percentile or the *first* quartile—called Q_1—is the point below which lie 25% of the scores (N). The 75th percentile or *third* quartile—called Q_3—is the point in the distribution below which lie 75% of the total number of scores (N). When we have these two points, the quartile deviation can be calculated by the formula:

$$Q = \frac{Q_3 - Q_1}{2} \tag{5}$$

(quartile deviation or Q calculated from a frequency distribution)

It is clear from the formula that to find Q we must first compute Q_1 and Q_3. These points are located in the same way as was the median, which is, of course, the 50th percentile or second quartile, Q_2. The only difference is that ¼ N is counted off to find Q_1 and ¾ N in order to find Q_3.

The formulas are:

$$Q_1 = l + i\left(\frac{\frac{N}{4} - cum\, f_l}{f_m}\right) \tag{6A}$$

and

$$Q_3 = l + i\left(\frac{\frac{3N}{4} - cum\, f_l}{f_m}\right) \tag{6B}$$

(the quartiles Q_1 and Q_3 from the frequency distribution) where

l = lower limit of the i upon which the quartile falls
i = the interval
$cum\, f_l$ = cumulated f up to the i containing the quartile wanted
f_m = f on the i which contains the quartile

Table 9 shows the computation of Q for the 60 AGCT scores tabulated into a frequency distribution in Table 1. To find Q_1, we must count off 25% of the f [or ¼ N], namely, 15 scores from the lower end of the distribution. From the cumulated f's it is clear that 14 scores complete the i (90-94) and take us to 94.5; and that Q_1 must lie on the interval (95-99). We know (see data in Table 9) that

l = 94.5, lower limit of i upon which Q_1 falls
¼ N = 15
$cum\, f_l$ = 14, sum of scores on i's below l
f_m = 10, frequency on the i containing Q_1
i = 5

From formula (6A)

$$Q_1 = 94.5 + 5\left(\frac{15\text{-}14}{10}\right)$$
$$= 95.0$$

TABLE 9

Calculation of Quartile Deviation (Q) from a Frequency Distribution. Data are 60 AGCT scores from Table 1.

Intervals	f	*cum* f
120-124	3	60
115-119	4	57
110-114	6	53
105-109	8	47
100-104	15	39
95-99	10	24
90-94	7	14
85-89	4	7
80-84	3	3
	$N = 60$	

$$\frac{N}{4} = 15; \qquad \frac{3N}{4} = 45$$

$$Q_1 = 94.5 + 5\left(\frac{15-14}{10}\right) = 95.00 \qquad \text{by formula (6A)}$$

$$Q_3 = 104.5 + 5\left(\frac{45-39}{8}\right) = 108.25 \qquad \text{by formula (6B)}$$

$$Q = \frac{108.25 - 95.00}{2} = \frac{13.25}{2} = 6.62$$

To find Q_3 we must count off $\frac{3}{4}N$ from the low end of the distribution or $\frac{1}{4}N$ from the high end. It is usually easier to count upward 75% of N than to count downward into an interval. Data from Table 9 are

$l = 104.5$, lower limit of the i upon which Q_3 lies
$\frac{3}{4}N = 45$
$cum\ f_l = 39$, sum of scores up to the i which contains Q_3
$f_m = 8$, f on the i containing Q_3
$i = 5$

From formula (6B)

$$Q_3 = 104.5 + 5\ \left(\frac{45 - 39}{8}\right)$$

$$= 108.25$$

Substituting in formula (5), we have that

$$Q = \frac{108.25 - 95.00}{2} = 6.62$$

Table 10 provides a second illustration of the computation of Q from a frequency distribution. The data are the 660 *IQ*'s of runaway boys tabulated in Table 3. The statistics to be put in formulas (6A) and (6B) will be found in Table 10.

To find Q_1:

$l = 59.5$, lower limit of the i upon which Q_1 falls

$\frac{1}{4}N = 165$

$cum\ f_l = 68$, $cum\ f$ on i's below l, namely 59.5

$f_m = 140$, f on i containing Q_1

$i = 10$

$$Q_1 = 59.5 + 10\left(\frac{165 - 68}{140}\right) = 66.43$$

To find Q_3:

$l = 79.5$, lower limit of i upon which Q_3 falls

$\frac{3}{4}N = 495$

$cum\ f_l = 392$, cumulated f on i's below l, (79.5)

$f_m = 139$, f on i containing Q_3

$i = 10$

By formula (6B)

$$Q_3 = 79.5 + 10\ \left(\frac{495 - 392}{139}\right)$$

$$= 86.91$$

and

$$Q = \frac{86.91 - 66.43}{2} = 10.24$$

TABLE 10

Calculation of Q from Frequency Distribution (the IQ's of 660 runaway boys). Data are from Table 3.

Intervals	*f*	*cum f*
110-119	14	660
100-109	37	646
90-99	78	609
80-89	139	531
70-79	184	392
60-69	140	208
50-59	60	68
40-49	6	8
30-39	2	2
	$N = 660$	

$$\frac{N}{4} = 165; \qquad \frac{3N}{4} = 495$$

$$Q_1 = 59.5 + 10\left(\frac{165 - 68}{140}\right) = 66.43 \qquad \text{By formula (6A)}$$

$$Q_3 = 79.5 + 10\left(\frac{495 - 392}{139}\right) = 86.91 \qquad \text{By formula (6B)}$$

$$Q = \frac{86.91 - 66.43}{2} = \frac{20.48}{2} = 10.24$$

The meaning of Q will not be apparent immediately to the student, owing in part to the necessity for mastering the many details of computation. The value of Q as a measure of variability lies in the fact that the two quartiles, Q_1 and Q_3, represent the 25% and the 75% points in the distribution. The distance between them gives the range of the middle 50%; and for this reason $(Q_3 - Q_1)$, the "interquartile range," is often called the "range of the middle 50." The interquartile range gives the boundaries within which falls the most typical part of the distribution; and Q, the semi-interquartile range, is ½ of this range $(Q_3 - Q_1)$. Hence, Q becomes an index of scatter or dispersion. In the normal probability curve, Q corresponds

to the "probable error" or *PE* (see p. 77) *Q* is an absolute,* not a relative measure of variability. We cannot, for instance, state that a *Q* of 6.62 or of 10.24 is large or small except in relation to some other *Q* computed from comparable data. When the means of two groups are not very different (boys and girls of the same age, or two sections of the same school grade, for example) their *Q*'s may be compared directly.

THE STANDARD DEVIATION OR SD

The standard deviation or *SD* is a measure of variability calculated around the mean. The *SD* is the most stable measure of variability and is customarily used in research problems and in those studies involving correlation. The *SD* is generally computed when the mean is the measure of central tendency, and the *Q* when the *Mdn* is the measure of central tendency. The usual symbol for the *SD* is the Greek letter σ (sigma).

Calculation of σ from Ungrouped Scores

Suppose that we have 5 scores as follows: 9, 8, 7, 6, 5, and wish to find their *SD*. Tabulating these scores we have

(1)	(2)	(3)
Scores (X)	$(X - M) = x$	x^2
9	2	4
8	1	1
7	0	0
6	−1	1
5	−2	4
$5\overline{)35}$		$\Sigma x^2 = 10$
$M = 7$		

* Absolute variability is variability expressed in the units of measurement used in the test: it is not relative to some other measure.

The M is simply 35/5 or 7. Now if we subtract the M from each individual score (X), we have in column (2) the x-deviations around the M of 7. These deviations $[(X - M) = x]$ are in order: 2, 1, 0, -1, and -2. Squaring each x we have in column (3) 4, 1, 0, 1 and 4, the sum being 10. The definition of σ when scores are ungrouped and each taken at face value is

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} \tag{7}$$

(SD computed from ungrouped scores) ✓

and in the present problem $\sigma = \sqrt{\frac{10}{5}} = 1.41$. The value of the SD as a measure of variability will be clearer when we consider the normal probability curve (p. 76). At this point we may note that squaring each x gives increased weight to extreme deviations and makes all deviations from M positive.

Calculation of σ from a Frequency Distribution

(1) THE SD BY DIRECT CALCULATION FROM MIDPOINTS AS SCORES

Table 11 shows the direct calculation of σ from two frequency distributions. The data in example (A) are the *IQ*'s of 660 runaway boys tabulated in Table 3. The procedure is the same as that used with ungrouped scores except for use of a midpoint and a column for f. First, the deviation of each interval midpoint (X) from M (see Table 5, p. 30) is found and tabulated in column (4) under x. (The midpoints, it will be remembered, represent *all* of the scores within the intervals.) Each deviation x is now multiplied by its corresponding f to give the fx in column (5). Multiplying x and fx entries on the same line in column (4) and column (5) gives column (6) the fx^2, and this is the column used in the calcu-

TABLE 11

Calculation of the SD (σ) by the Direct Method

A. Data from Table 3, p. 18, are the *IQ*'s of 660 runaway boys

(1) *IQ-intervals*	(2) *midpoint*	(3) *f*	(4) *x*	(5) *fx*	(6) *fx*²
110-119	114.5	14	37.33	522.62	19509.40
100-109	104.5	37	27.33	1011.21	27636.37
90-99	94.5	78	17.33	1351.74	23425.65
80-89	84.5	139	7.33	1018.87	7468.32
70-79	74.5	184	−2.67	−491.28	1311.72
60-69	64.5	140	−12.67	−1773.80	22474.05
50-59	54.5	60	−22.67	−1360.20	30835.73
40-49	44.5	6	−32.67	−196.02	6403.97
30-39	34.5	2	−42.67	−85.34	3641.46
		$N = 660$			142706.67

$M = 77.17$ (Table 5, p. 30)

$$SD \text{ or } \sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{142706.67}{660}} = \sqrt{216.2222} = 14.70$$

B. Data from Table 3, p. 19, give the number of pupils in 43 classrooms in three elementary schools

(1) (*No. of pupils*) *Intervals*	(2) *midpoint*	(3) *f*	(4) *x*	(5) *fx*	(6) *fx*²
36-38	37	2	7.40	14.80	109.52
33-35	34	5	4.40	22.00	96.80
30-32	31	17	1.40	23.80	33.32
27-29	28	10	−1.60	−16.00	25.60
24-26	25	8	−4.60	−36.80	169.28
21-23	22	1	−7.60	−7.60	57.76
		$N = 43$			492.28

$M = 29.60$ (Table 3, p. 19)

$$SD \text{ or } \sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{492.28}{43}} = \sqrt{11.4484} = 3.38$$

lation of the SD. The first fx^2 entry of 19509.40 is found from 37.33 × 522.62, i.e., $x \times fx = fx^2$. Since all negative x entries are squared, all fx^2 are positive. The sum of the fx^2 is 142706.67. Dividing this number by 660 (N) and extracting the square root we get 14.70 as the SD. The formula is

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} \tag{8}$$

(SD computed from a frequency distribution)

In the second example in Table 11, exactly the same procedure is followed as above. The M of 29.60 is taken from Table 3 (p. 19) and in column (4) the deviation of the midpoint of each i from 29.60 is recorded. The fx in column (5) are found by multiplying corresponding entries in columns (3) and (4). And the fx^2 in column (6) are products of corresponding entries in columns (4) and (5). The SD by formula (8) is

$$SD = \sqrt{\frac{492.28}{43}} = \sqrt{11.4484} = 3.38$$

This method of calculating SD by finding deviations of midpoints from the M is straightforward and relatively simple; but it is lengthy and involves much multiplication of decimals. This procedure is most useful when (1) scores are ungrouped and are few in number, and (2) when the M is an integer so that the x's are whole numbers. For the routine computation of the SD from a frequency distribution, the method in which M is first assumed is to be recommended. It is far quicker and more efficient than the direct method.

(2) THE SD BY THE ASSUMED MEAN METHOD

On page 29 a method was given for computing M from a frequency distribution by first assuming a mean (AM) and

then applying a correction to give the actual M. The assumed mean method has decided advantages over the direct method when the SD is to be computed as well as the M; and in the calculation of the coefficient of correlation from a diagram it is well-nigh indispensable (p. 113).

The plan of computation is shown in Table 12. The two examples are the same as those used in Table 11 to illustrate the direct calculation of the SD around the interval midpoints. Tables 11 and 12 permit, therefore, a direct comparison of the two procedures. The following steps outline the method:

(a) First assume M preferably on the i having the largest f and as close to the center of the distribution as possible (p. 31).

(b) In column (3) headed x' * list the deviation of each i midpoint from the AM in *units-of-interval.* In example (A), Table 12, the entries are 1, 2, 3, 4 up from 0; and −1, −2, −3, −4 down from 0; the AM is 74.5. In example (B) the x' entries are 1, 2 up from 0, and −1, −2, −3 down from 0; the AM is 31.00.

(c) In column (4) tabulate the fx' found by multiplying each x' by its appropriate f. Sum the plus and minus fx' separately and compute the correction (c) by the formula $c = \frac{\Sigma fx'}{N}$ (algebraic) (see p. 32). In example (A) the c is 176/660 or .2667; in example (B), the c is −20/43 or −.4651. In example (A), c^2 is .0711 and in example (B) c^2 is .2163.

(d) In column (5) tabulate the fx'^2. These entries are found most readily by multiplying the corresponding x' and fx' entries. In the first example the sum of the fx'^2 is 1474 and in the second example $\Sigma fx'^2$ is 64. All of the fx'^2 entries are positive, since the x' have all been squared.

* $X - AM = x'$, just as $X - M = x$. See p. 31.

(e) The formula for finding the *SD* by the assumed mean method is

$$\sigma = i\sqrt{\frac{\Sigma fx'^2}{N} - c^2} \tag{9}$$

(σ by the assumed mean method)

in which

$\Sigma fx'^2$ = sum of the squared deviations around the *AM*
i = the interval
N = the number of cases
c = the correction in units-of-interval

TABLE 12

Calculation of SD (σ) by the Assumed Mean Method

A. Data from Table 3, p. 18: *IQ*'s of 660 runaway boys

(1) *IQ-intervals*	(2) f	(3) x'	(4) fx'		(5) fx'^2
110-119	14	4	56		224
100-109	37	3	111		333
90-99	78	2	156		312
80-89	139	1	139		139
70-79 (74.5)	184	0	0	462	0
60-69	140	−1	−140		140
50-59	60	−2	−120		240
40-49	6	−3	−18		54
30-39	2	−4	−8		32
	$N = 660$			−286	1474
				176	

Assumed $M = 74.5$ $\quad c = \frac{\Sigma fx'}{N} = \frac{176}{660} = .2667 \quad c^2 = .0711$

$$\sigma = i\sqrt{\frac{\Sigma fx'^2}{N} - c^2} = 10\sqrt{\frac{1474}{660} - .0711} = 10\sqrt{2.1622} = 14.70$$

B. Data from Table 3, p. 19: number of pupils in 43 classrooms

(1) *Intervals (number pupils)*	(2) f	(3) x'	(4) fx'		(5) fx'^2
36-38	2	2	4		8
33-35	5	1	5		5
30-32 (31)	17	0	0	9	0
27-29	10	−1	−10		10
24-26	8	−2	−16		32
21-23	1	−3	−3		9
	N = 43			−29	64
				−20	

$$AM = 31.00 \qquad c = \frac{\Sigma fx'}{N} = \frac{-20}{43} = -.4651 \qquad c^2 = .2163$$

$$\sigma = i\sqrt{\frac{\Sigma fx'^2}{N} - c^2} = 3\sqrt{\frac{64}{43} - .2163} = 3\sqrt{1.2721} = 3.38$$

In example (A) the *SD* is $10\sqrt{\frac{1474}{660} - .0711}$ or 14.70; and in example (B) the σ is $3\sqrt{\frac{64}{43} - .2163}$ or 3.38. These results check exactly with those found by the method of direct calculation from midpoints. Even with the computation of c and c^2, the *AM* method is easier and requires less time and arithmetic than the straightforward procedure.

A word may be said about formula (9). When the plus and minus fx' balance exactly, $c = 0$ and the prime (′) is no longer needed for x. Formula (9) now reduces to (8)—deviations are from the actual mean. The computation of c^2 should be taken to four decimal places, in order that the σ may be accurate to at least two decimals when the square root is taken.

(3) CALCULATION OF THE SD FROM RAW OR OBTAINED SCORES

It is sometimes desirable to compute *SD* directly from raw or obtained scores without first getting deviations from *M* or tabulating the data into a frequency distribution. This method is especially valuable when *N* is small, the scores ungrouped, and a calculating machine is available. Table 13 provides two examples. In the first, *M* as given by formula (1) is 21.50 (258/12). In the second column each score (X) has been squared, and the sum of these squares (X^2) when put into the formula below gives the *SD*:

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - M^2} \tag{10}$$

(σ from raw or obtained scores)

Substituting for $\Sigma X^2 = 5894$, for $N = 12$ and $M^2 = 462.25$, we have that the *SD* is 5.38.

Formula (10) follows directly from formula (9) in which the σ was computed by the assumed mean method. If the *AM* is taken at 0, each score (X) becomes at once a deviation (x') from this *AM* of 0; that is, each score is unchanged. The correction (c) is always the difference between the *M* and the *AM*. When the *AM* is 0, the c in the equation ($M - AM = c$) becomes simply the *M*. And, of course, $c^2 = M^2$. An alternate formula to (10) is obtained by replacing M^2 by $\frac{(\Sigma X)^2}{N^2}$. We then have

$$\sigma = \frac{\sqrt{N\Sigma X^2 - (\Sigma X)^2}}{N} \tag{11}$$

(alternate formula for σ when computed from obtained scores)

Formula (11) is used in example (B) Table 13 to illustrate the calculation of the *SD*. For convenience in calculation, the 10 scores on the personality inventory shown in Table 13, ex-

ample (B), have first been reduced by subtracting 90 (the lowest score) from each.* Each $(X - 90)$ or "new" X is then squared and added to give $\Sigma X^2 = 2687$. The sum of the reduced scores $(X - 90)$ is 141. Substituting for ΣX^2 and for $(\Sigma X)^2$ in the formula, we have

$$\sigma = \frac{\sqrt{10 \times 2687 - (141)^2}}{10} = \frac{\sqrt{6989}}{10} = 8.36$$

TABLE 13

Computation of SD from Obtained Scores

A. The data are 12 ungrouped scores on a reading test

Scores (X)	X^2
19	361
28	784
25	625
30	900
16	256
12	144
21	441
18	324
17	289
29	841
23	529
20	400
258	5894

$$M = \frac{\Sigma X}{N} = \frac{258}{12} = 21.50 \qquad \text{by formula (1)}$$

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - M^2} = \sqrt{\frac{5894}{12} - (21.5)^2} = \sqrt{28.92} = 5.38 \qquad \text{by (10)}$$

* Subtraction of a constant from each score moves all of the scores down by that amount, but does not change the relative position of the scores or their variability. Subtracting 90 from each score in example (B) reduces the M by just 90—from 104.1 to 14.1; but the variability of the scores (the σ) remains unchanged.

B. Data are 10 ungrouped scores on a personality inventory

Scores (X)	$(X-90)$ or "reduced X"	X^2
110	20	400
102	12	144
90	0	0
100	10	100
112	22	484
115	25	625
95	5	25
110	20	400
112	22	484
95	5	25
$M = 104.1$	$(\Sigma X) = 141$	$\Sigma X^2 = 2687$

$$\sigma = \frac{\sqrt{N\Sigma X^2 - (\Sigma X)^2}}{N}$$

$$= \frac{\sqrt{26870 - 19881}}{10} = 8.36 \qquad \text{by (11)}$$

WHEN TO USE THE DIFFERENT MEASURES OF VARIABILITY

In general we

Use the Range
(1) when the data are scattered or scanty and only a general guide as to variability is wanted
(2) when a single rough measure of total spread is desired

Use the Q
(1) for a quick inspectional measure of variability
(2) when the *Mdn* is the measure of central tendency
(3) when interest centers in the middle of the distribution

Use the SD

(1) when extreme deviations should have proportionately greater weight upon the measure of variability
(2) when r's are later to be computed (see p. 113)
(3) when the measure of variability of highest reliability is wanted

Note on the Average Deviation (AD) or the Mean Deviation (MD)

While the AD is rarely used today, it is found in older experimental literature and has the virtue of simplicity. This measure of variability is defined simply as the sum of the deviations (x) around the mean—without regard to sign*—divided by N. The formula is

$$AD \text{ or } MD = \left|\frac{\Sigma x}{N}\right| \text{(i.e., without regard to sign)} \qquad (12)$$

(the AD or MD calculated around the mean)

In the example of the 5 ungrouped scores on page 51 the AD is simply $\frac{6}{5}$ or 1.20. All of the x's are treated as though they were positive in sign. In Table 11, example A, the AD is 11.84: 7811.08 [the sum of the fx's in column (5)] divided by 660. The formula here differs from (12) in that fx is substituted for x. In Table 11, example B, the AD is 121/43 or 2.81. In both of these examples the fx are added without regard to sign.

The AD is always larger than the Q and smaller than the SD. This relation provides a rough check upon its accuracy.

* The sum of the deviations taken around the mean always equals zero. When added without regard to sign, the average or mean of the deviations gives the spread of the scores around the M in absolute amount.

5.

PERCENTILES AND PERCENTILE RANKS

Percentiles are points found by counting off a given percentage of N when a set of scores has been arranged in order of size. We have already had occasion to use percentiles in computing the *Mdn* and the Q from a frequency distribution. The median is the point found by counting 50% of the way into the distribution; hence, the *Mdn* is the 50th percentile or P_{50} on a scale of 100 units. The first quartile, as we know (p. 46), is found by counting off 25%, and the third by counting off 75% of N, the total frequency. Q_1 is the 25th percentile or P_{25}, and Q_3 is the 75th percentile or P_{75}.

When we calculate percentiles, the score distribution is thought of as stretched out or compressed into a scale of 100 units, called the *percentile scale*. We may count any distance up to 100% along this scale. If, for example, we wish to cut off the lowest 20% of a group, we must count off 20% of the scores. This will locate P_{20} or the 20th percentile point—the point below which lies 20% of the distribution. Other percentiles are counted off in the same way.

The percentile scale may be used in another way. Suppose that we wish to compare the scores made by John, a sixth grader, on two tests. If we know that John's score on an arith-

metic test is 24 and that his score on a reading test is 63, we cannot tell from these scores as they stand how well the boy has done on either test. However, if we know that a score of 24 in arithmetic has a *percentile rank* (*PR*) of 60, and that a score of 63 in reading has a *PR* of 50, we can say at once that John is better in arithmetic than he is in reading, and that he did fairly well in both tests. A *PR* of 60 on the arithmetic test means that 60% of sixth graders fell below John in arithmetic achievement; and a *PR* of 50 on the reading test means that 50% of sixth graders fell below him in reading. John is above the *Mdn* in arithmetic and just on the *Mdn* in reading. Said differently, on a scale of 100 units, John ranks 60 from the bottom in arithmetic and 50 from the bottom in reading.

These examples point up the distinction between percentile and percentile rank. Percentiles are *points* in the distribution found by counting off any given per cent of N: for example, to find P_{15}, P_{42}, P_{81}, we count off 15%, 42% and 81% from the low end of the distribution. The *PR* of a score, on the other hand, is that proportion of N which falls *below* the midpoint of the given score. If a score of 82 on a history examination has a *PR* of 58, this means that 58% of those taking the test scored *below* the score of 82 (remember that 82 is the midpoint of the score-interval 81.5 – 82.5).

Reading Percentiles and Percentile Ranks from an Ogive

Figure 8 shows the cumulative percentile curve or *ogive* plotted from the data in Table 1, p. 14. These 60 scores have been cumulated or added progressively from the bottom of the distribution up as shown in Table 14, and the cumulated scores have been expressed as per cents of N (by dividing by 60) in the column headed "*cum % f*." In Figure 8 the *cum % f* have been plotted on the Y- or vertical axis against scores on the X-axis to give the ogive. Since the progressive adding of

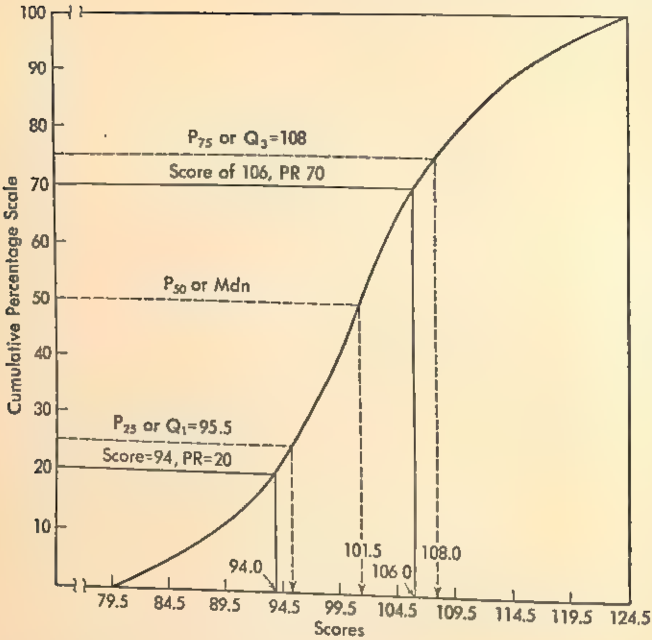


Figure 8

*f*s carries us to the upper limits of the intervals, in each instance the *cum % f*s have been plotted just above the *upper* limits of the *i*'s:—5% above 84.5, 11.7% above 89.5, 23.3% above 94.5 and so on.

Both percentiles and percentile ranks can easily be read from the ogive. For example, to find P_{50}, the *Mdn*, we go from 50 on the *Y*-scale across to the curve, and drop a perpendicular to the *X*-axis to locate the *Mdn* approximately at 101.5. The computed value of the *Mdn* is also 101.50. Graphic location of percentiles is not always as precise as this, but is good enough for many purposes. Accuracy of location will depend upon the size of *N*, the smoothness of the plot, the fineness of the interval-units, and the care taken in drawing the graph. The values of Q_1 and Q_3 when read from the ogive, are ap-

proximately 95.5 and 108.0; the computed values shown in Table 14 are 95.0 and 108.25.

TABLE 14

Data from Table 1, p. 14, 60 AGCT scores

Intervals	*f*	*cum f*	*cum % f*
120-124	3	60	100.0
115-119	4	57	95.0
110-114	6	53	88.4
105-109	8	47	78.3
100-104	15	39	65.0
95-99	10	24	40.0
90-94	7	14	23.3
85-89	4	7	11.7
80-84	3	3	5.0
	$N = 60$		

$N/4 = 15 \quad 3N/4 = 45$

$$Q_1 = 94.5 + 5\left[\frac{15 - 14}{10}\right] = 95.0$$ formula (6A)

$$Q_3 = 104.5 + 5\left[\frac{45 - 39}{8}\right] = 108.25$$ formula (6B)

To find the *PR* (percentile rank) of any score, start with the score on the *X*-axis, go up to the curve and across to read the *PR* on the *Y*-axis (the percentile scale). A score of 106 (midpoint of score-interval 105.5-106.5) has a *PR* of approximately 70: 70% of the frequency (*N*) lies below this point. A score of 94 has a *PR* of approximately 20, as read from the ogive.

The ogive in Figure 9 pictures the *cum % f* distribution of scores made by 1000 adults on the SRA* Non-Verbal Test of General Intelligence. This ogive constitutes a compact set

* Science Research Associates. See Examiner's manual, SRA verbal and non-verbal forms, Chicago: 1947.

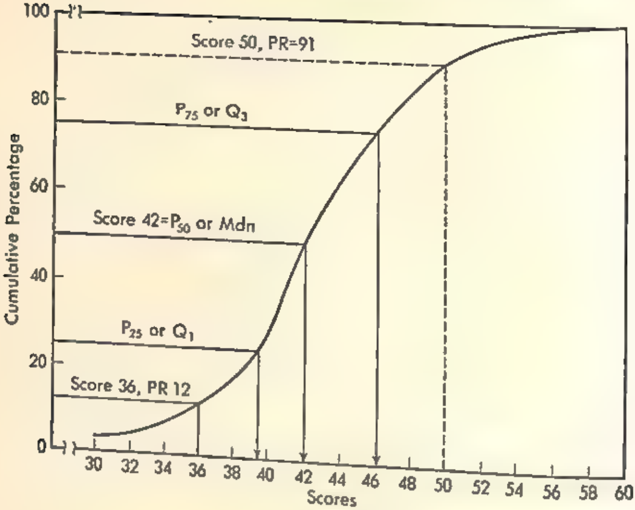


Figure 9

of *norms* for this test. If a candidate achieves a score of 50 his *PR* is approximately 91; a score of 36, his *PR* is 12; a score of 42, his *PR* is 50. Note also that the *Mdn* or P_{50} is approximately 42, that Q_1 is approximately 39.5, that Q_3 is approximately 46.5. In reading *PR*'s it should always be remembered that the percentage counted off is the fraction of *N* which lies *below* the midpoint of the given score.

The Computation of Percentiles from a Frequency Distribution

When percentiles read from the ogive are not judged to be sufficiently accurate, it is possible to compute them directly from a frequency distribution using the same procedure as that followed in finding the *Mdn* and the quartiles. The formula, with appropriate changes, is the same as that for the *Mdn*:

$$P_p = l + i\left(\frac{pN - F}{f_p}\right) \tag{13}$$

(computing percentiles from a frequency distribution, counting from the low end of the distribution)

in which

p = percentage wanted
pN = % of N to be counted off to reach P_p
l = lower limit of i containing P_p
F = sum of f's on i's below l
f_p = f on i upon which P_p falls
i = interval

To illustrate, if we should want P_{62} from Table 14, the data are

p = 62%
pN = 37.20, i.e., 62% of 60
l = 99.5, lower limit of (100-104), the interval containing P_{62}
F = 24, number of f's below 99.5
f_p = 15, f on i (100-104) upon which P_{62} falls
i = 5

Substituting in formula (13)

$$P_{62} = 99.5 + 5\left(\frac{37.20 - 24}{15}\right) = 103.9$$

To take another illustration, suppose that we want P_{83} from Table 14. Then

p = 83%
pN = 49.80 or 83% of 60
l = 109.5, lower limit of (110-114)
F = 47
f_p = 6
i = 5

and $$P_{83} = 109.5 + 5\left(\frac{49.80 - 47}{6}\right) = 111.83$$

These two percentiles as read from the ogive are 104 and 111.5, values which check closely with the more accurately calculated points.

Computing PR's from a Frequency Distribution

The *PR* of a score may also be computed directly from the frequency distribution when a more accurate value than that estimated from the ogive is desired. Suppose that a man has achieved an AGCT score of 117 and that we want to find his *PR* from the frequency distribution in Table 14. The procedure is as follows:

(1) Count off scores up to the beginning of the i which contains the given score. Score 117 falls on interval (115-119) and there are 53 f's (cumulated scores) up to 114.5, lower limit of (115-119).

(2) Divide the f on the interval by the length of i. Here we divide the 4 f's on (115-119) by 5 (the interval) to yield .8 score *per unit-of-interval.*

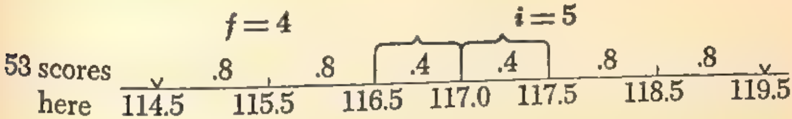
(3) Find the distance of the given score (namely, 117) from the lower limit of the interval which contains it. In the present case, $117 - 114.5 = 2.5$; and 2.5 is the distance of score 117 from the lower limit of the interval (115-119).

(4) Multiply the distance of the score from the beginning of the interval, by the number of scores per unit-of-interval. In the present case, $2.5 \times .8$ gives 2.0.

(5) Add the result obtained in (4) above to 53 to give the frequency up to 117.0, midpoint of the score-interval (116.5-117.5). This gives $53 + 2$ or 55, as that part of $N(60)$ *below* score 117.

(6) Divide the number of scores falling below 117 by N to give the percentage of N below 117. In our problem,

55/60 = .917; and 92 is, therefore, the *PR* of score 117. A diagram will make the computations clearer:



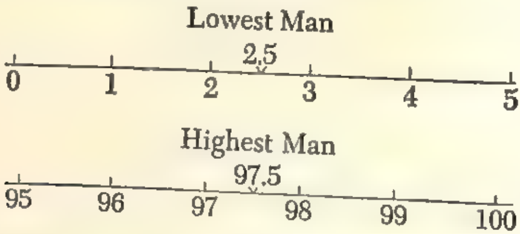
Fifty-three scores lie below 114.5. Prorating the 4 scores on (115-119) over the interval of 5 units, we have .8 score per unit-of-interval. The score of 117 lies .8 + .8 + .4 or 2.0 from 114.5, the beginning of the interval. Hence, score 117 is 53 + 2 or 55/60 of the way into the distribution.

The *PR* read from the ogive in Figure 8 checks almost exactly with 92, the computed *PR* of 117. Direct calculation of *PR*'s from the frequency distribution is sometimes necessary. But the graphic determination of *PR*'s is generally to be recommended as being faster and for practical purposes just as useful. Note again that the *PR* of a score is the per cent of *N* lying *below* the midpoint of the score.

Computing PR's from Ranked Data

It is sometimes convenient to put the members of a group in order of merit for some personality trait, or to rank them in 1-2-3 order on the basis of scores received on an examination. Standing in orders of this sort when expressed in terms of *PR*'s facilitates comparison and renders interpretation easier. To illustrate, suppose that 20 men have been ranked in order of merit after an interview designed to assay extent of experience and suitability for employment. Each man is given a *PR* in the following way. As there are exactly 20 men, each may be thought of as occupying 100/20 or 5 units along the percentile scale. The lowest ranking man then has a *PR* of 2.5

and the highest a *PR* of 97.5: each man is given the *PR* corresponding to the midpoint of the 5-unit interval which he occupies. The diagram below shows this:



For each rank, a man's *PR* is the midpoint of the 5-unit interval allotted to his position on the scale of 100 units. A formula which converts orders of merit into *PR*'s is

$$PR = 100 - \frac{(100R - 50)}{N} \tag{14}$$

(*PR*'s for individuals or things arranged in order of merit) in which

PR = percentile rank
R = original rank or position in the list
N = size of sample

By means of formula (14) the man who ranks second from the top in a group of 20 has a *PR* of $100 - \frac{(100 \times 2 - 50)}{20}$ or 92.5; the man who ranks third, 87.5, etc. The *PR* formula is most useful when *N* is odd and is fairly large. What, for example, is the *PR* of a girl who ranks twelfth from the top in an English class of 37? By formula (14), her *PR* is $100 - \frac{(100 \times 12 - 50)}{37}$ or 69. This *PR* may be compared directly with other *PR*'s since they are expressed in terms of the same scale of 100 units. Suppose that the girl who ranks twelfth in the English class of 37 also ranks twelfth in a mathematics

class of 22. How do these two "standings" compare? In the second case, the girl's *PR* is $100 - \frac{(100 \times 12 - 50)}{22}$ or 48. It seems clear that this student is better in English (*PR* of 69) than she is in mathematics (*PR* of 48) in spite of the fact that she is twelfth from the top in both classes.

6.

THE NORMAL PROBABILITY DISTRIBUTION AND THE NORMAL CURVE

The frequency polygons and histograms shown in Figures 1 and 3 are alike in the following respects: scores tend to be bunched or concentrated at the center of the scale, and to taper off gradually and fairly regularly to the right and left of a central high point. This means that relatively few persons score at the extremes of the distribution, that is, either very high or very low, and that most of the group achieve intermediate scores. If we divide the area of the polygon in Figure 1 into two parts, by cutting along a perpendicular line drawn through the peak or high point, the right section of the figure will match almost exactly the left section: the two "halves" will be virtually equal. Many frequency polygons plotted from all sorts of data approach in general form the symmetrical curve shown in Figure 10. This *normal probability curve* or more simply *normal curve*, as it is usually called, is widely used in statistical theory and practice. It serves as a model for many distributions of scores; it enables us to make predictions and generalizations; and it is almost indispensable in mental test construction and in test theory. This chapter will be concerned with the main characteristics and applications of the normal curve.

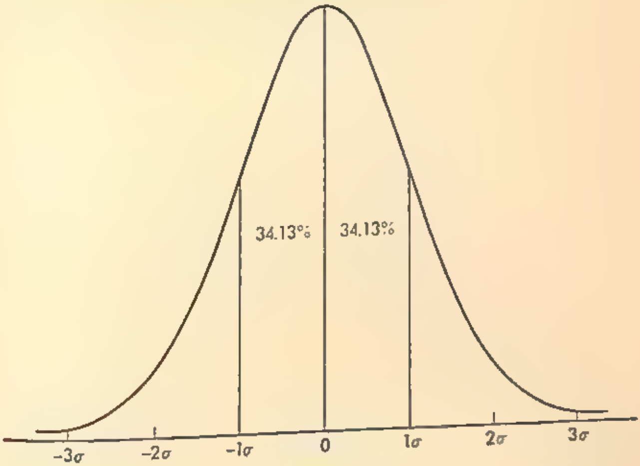


Figure 10

The Normal Distribution

The normal curve in Figure 10 is a smooth, bell-shaped frequency polygon which represents graphically the normal distribution. The normal distribution is not an actual distribution of test scores or of other measures. Instead, it is a mathematical model—a theoretical distribution—the area (N) of which is taken to be infinitely large. The normal distribution, and its frequency polygon, the normal probability curve, may be thought of as arising from the operation of a very large number of elementary factors. These factors are conceived to be similar, equal and independent—as likely to be present as absent. How such factors operate to produce a normal distribution may be demonstrated most simply perhaps in the following way. Suppose that six coins are tossed and the number of heads and tails that appear are counted. It is evident that we may get four heads and two tails, or three heads and three tails, or some other combination, such as six heads and no tails. A coin has only two faces so that a

head or tail is equally probable, and the probability (p) of either face is ½. If we let $p =$ probability of a head and $q =$ the probability of a tail, then $(p + q)$ equals $(\frac{1}{2} + \frac{1}{2})$ or 1.00. Moreover, if we write $(p + q)^6$ the expansion of this binomial * will give the *expected* occurrences of the 64 combinations of heads and tails when six coins are tossed. Thus, we have

$$(p + q)^6 = p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6$$

The first of these terms, p^6, gives the frequency of occurrence (namely, 1) of six heads and no tails (see Table 15, below); the second term, $6p^5q$, gives the frequency of five heads and one tail (i.e., 6); the third term, $15p^4q^2$, gives the frequency of occurrence of four heads and two tails (i.e., 15), and so on down to q^6, which gives the frequency of no heads and six tails. The binomial expansion will have more meaning for present purposes when written as a frequency distribution in which the number of heads is taken as the "score":

TABLE 15

Frequency Distribution Written from the Expansion of the Binomial $(p + q)^6$

Score No. of heads	*f*	*Probability of various numbers of heads*	
6	1	Probability of 6 heads:	1/64
5	6	Probability of 5 heads:	6/64
4	15	Probability of 4 heads:	15/64
3	20	Probability of 3 heads:	20/64
2	15	Probability of 2 heads:	15/64
1	6	Probability of 1 head:	6/64
0	1	Probability of 0 heads:	1/64
	$N = 64$		

* Binomial = an expression containing two terms. Rules for expanding a binomial will be found in any elementary algebra.

The probability or expectation of any number of heads from 6 to 0 may be found by dividing the appropriate f by N. Thus, the probability of all six coins falling heads is 1 in 64 since there are 64 possible combinations of heads and tails but only one in which all coins are heads. The highest probability, namely, 20/64 is for three heads and three tails. There are 6 chances in 64 that the coins will fall one head and five tails, and only 1 chance in 64 that no heads will appear (q^6)—that all coins will show tails. A frequency polygon and a histogram describing the frequency distribution in Table 15 is shown in Figure 11 below.

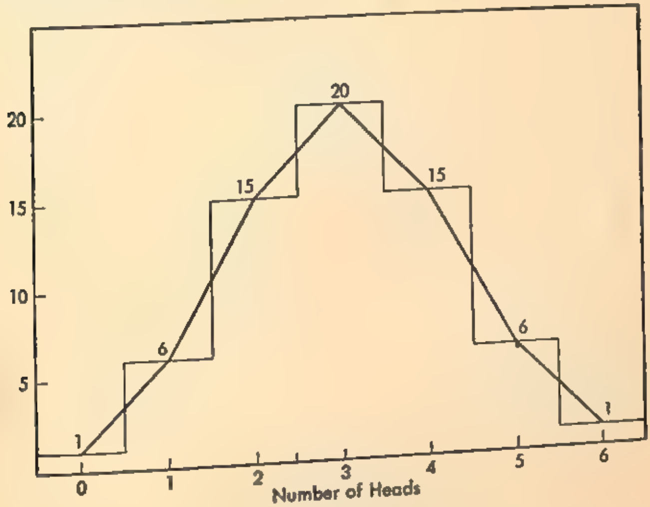


Figure 11

The frequency distribution in Table 15 and the frequency polygon in Figure 11 give the theoretical expectation of the occurrence of certain events (in the present case, combinations of heads and tails) when six coins are operating to produce a "score." When the number of operating factors is very large indeed, the frequency polygon in Figure 11 becomes

the smooth normal probability curve shown in Figure 10. The reason why many frequency distributions of test scores resemble closely those distributions obtained by tossing coins may be that many are actually probability distributions. The appearance on a coin of a head or a tail is determined by a large number of small ("chance") influences as likely to work one way as another. The twist given the coin, its weight, the kind of surface it falls upon, and other circumstances—all these may be important. By analogy, the presence or absence of any one of the undoubtedly large number of genetic factors that determine the shape of a man's head, or his intelligence, or his mechanical ability, may depend upon a host of influences the combination of which is governed by "chance." The fact that many distributions of obtained scores resemble the normal distribution does not, however, force us to the conclusion that distributions of mental traits are always or even usually "chance" arrangements. This is an interesting speculation, to be sure, but the genetic factors that determine musical aptitude or intelligence, for example, are too little known to warrant the assumption that they behave like coin combinations. The widespread use of the normal distribution is justified by the fact that it often fits the facts better—provides a better model—than other mathematical distributions.

Table of the Normal Distribution

Table I in the Appendix gives the fractions of area (N) under the normal probability curve measured from the M in units of standard deviation (σ). Since the two halves of the normal curve are exactly alike, the M, the Mdn, and the $Mode$ all fall at the same point and are equal. The total area of the curve is taken to be 1.00 or 100% so that entries in the body of Table I may be read as percentages. The first column on the left, headed x/σ, gives the distance along the base line from

M measured in units of σ. To find, for example, how much of the area lies between the M and 1.00σ, read down the x/σ column to 1.00 and in the next column under .00 take the entry 34.13. This figure means that 34.13% of the area of the normal distribution lies between the M and 1.00σ. To find how much of the area lies between the M and 1.65σ, read down the x/σ column to 1.6, and in the column headed .05 take the entry 45.05%.

The normal distribution is bilaterally symmetrical around the M, so that entries in Table I apply to the left as well as to the right half of the curve. If we go out -1.54σ from M, that is, to the *left* of M, we read from Table I that 43.83% of the area lies between M and this point. Between the M and ± 1.00 (x/σ measured in both directions from the M) lies 68.26% of the total area. An x/σ of 2.00 includes 47.72% of the area in the *right* half of the curve. Hence, $\pm 2.00\sigma$ measured out from the M (in both directions) takes in slightly more than 95% (47.72% $\times$ 2) of the area (N).

Entries in the body of Table I give the per cents of area between the M and various points measured out from the M in terms of σ. If it happens that a certain percentage of area is given, we may reverse the process and find how far we need to go out from M (in units of σ) in order to include the given proportion. Suppose, for example, that we wish to find how far we must go along the base line to include the 25% of area just *below* the M. From the body of the table, we take the entry 24.86% as our closest approximation to 25%. This per cent lies in the column headed .07 opposite .6 in the x/σ column and gives $-.67\sigma$ * as the distance we must go to take in the 25% just below the M. The middle 50% of the area lies between $\pm .67\sigma$; and the Q, called the PE in the normal curve, is equal to $.67\sigma$.

For all practical purposes, the normal curve may be said to

* By interpolation this value is found more exactly to be $-.6745\sigma$.

end at -3.00σ and $+3.00\sigma$, although mathematically the curve reaches the base line at infinity in both directions—to right and left. From Table I we read that 49.87×2 or 99.74% of the area of the normal curve lies between $\pm 3.00\sigma$. In cutting off the curve at these two points, therefore, we lose .26 of 1% of N, a negligible amount except in very large samples.

Applications of the Normal Probability Curve

A number of problems arise in experimental work which may be solved quite readily if we are justified in taking the normal curve as our model. Several of the more common applications of the normal curve will be treated in this section.

(1) TO FIND THE NUMBER OF SCORES WITHIN CERTAIN LIMITS IN A FREQUENCY DISTRIBUTION

Example (1): For children in general, the mean *IQ* on the Stanford-Binet intelligence test is 100 and the *SD* is 16. Let us suppose that all children of 70 *IQ* and below are to be taken from the regular classes and sent to special schools. If we assume the distribution of intelligence to be essentially normal, (a) How many children would be taken from the regular classes? (b) How many children would we expect to find within the *IQ* limits 90-110, inclusive? (c) What point in the *IQ* distribution marks off the *highest* 10% with respect to intelligence? (d) The *lowest* 20%?

(a) The upper limit of 70 *IQ* (namely, 70.5) lies at a distance of 29.5 *IQ* points from the *M* of 100: or, $70.5 - 100 = -29.5$. Dividing this deviation from the *M* by σ, we have $x/\sigma = \frac{-29.5}{16} = -1.84$ as the distance that 70.5 lies to the *left* of the *M*. (See Fig. 12.) From Table I, we find that 46.71% of a normal distribution lies between *M* and -1.84σ; and ac-

cordingly, 3.29% of the distribution lies to the *left* of -1.84σ. It follows, therefore, that about 3% of school children can be expected to have *IQ's* of 70 and below, and must be sent to special schools. It will be clear from Figure 12 why we must take the upper limit of 70 *IQ* (namely, 70.5), in order to include 70 *IQ* in the group to be sent to special schools. Note that about 3% of school children may also be expected to have

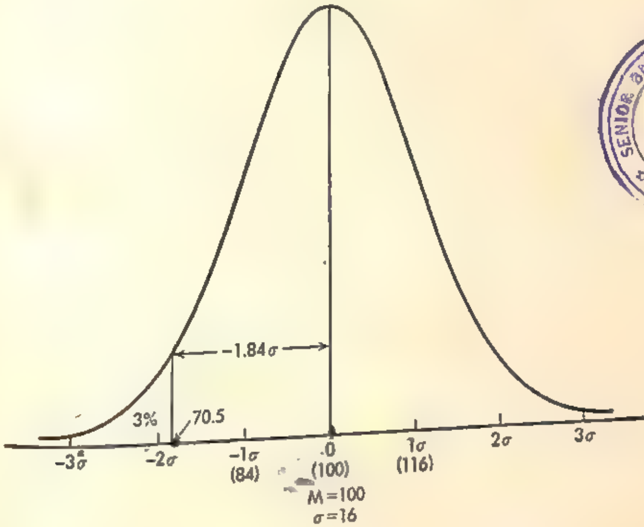


Figure 12

IQ's of 130 and above, to fall at the right extreme (good end) of the curve.

(b) How many unselected children (that is, children in general) would we expect to find within the *IQ* range 90-110 inclusive? The lower limit of 90 *IQ* is 89.5, and this point is 10.5 *below* the *M* of 100; also the upper limit of 110 *IQ* is 110.5 which lies 10.5 *above* the *M* of 100. Dividing the common deviation of 10.5 by the *SD* $(x/16)$ we have that these points lie $\pm 10.5/16$ or $\pm .66\sigma$ to the right and left of the *M*. From

Table I it is clear that roughly 49% (24.54×2) of school children in general can be expected to possess *IQ*'s from 90 to 110 inclusive. (See Fig. 13.)

(c) What point in the *IQ* distribution marks off the highest 10% with respect to intelligence? The highest 10% in a normal distribution is just 40% from the *M*; and from Table I we find that we must go out 1.28σ from the *M* in order to include the

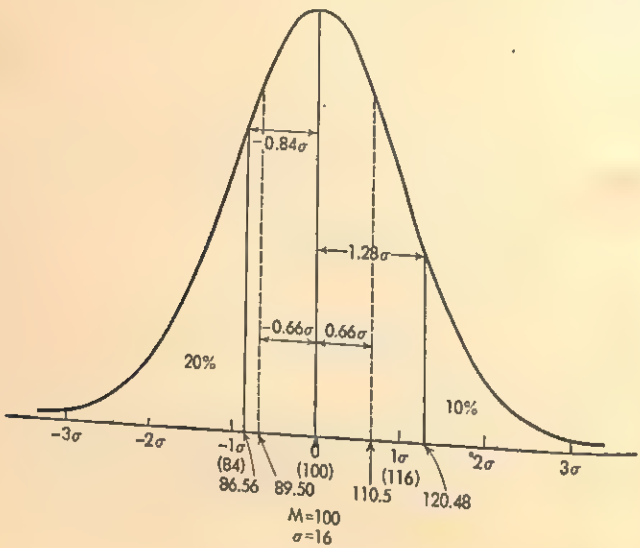


Figure 13

40% of area just above the *M*. The *SD* of the *IQ* distribution is given as 16. Hence, we must go 1.28×16 or 20.48 *IQ*-units above 100 to reach 120.48—the *point* in the *IQ* distribution above which lies 10% of the distribution. (See Fig. 13.)

(d) What point in the *IQ* distribution marks off the lowest 20%? The lowest 20% in a normal distribution is just 30% to the left of the *M*. From Table I it is clear that we must go out

to $-.84\sigma$ to include the 30% just below the M of 100. The σ of the IQ distribution is 16. Hence, we must go out $-.84 \times 16$ or -13.44 IQ points from 100 to reach 86.56, the *point* in the IQ distribution which marks off the lowest 20% of the distribution. (See Fig. 13.)

Example (2): Given a frequency distribution of scores on an Educational Achievement Test with M of 162 and SD of 30. On the assumption of normality in the distribution, (a) What limits mark off the middle 50% of scores? (b) the middle 75%? (c) What percentage of the distribution falls below the score 112?

(a) To include the 25% of N just above the M, we need to go out a distance of $.67\sigma$ * (Table I). And to include the 25% just below the M we must, of course, go out $-.67\sigma$. Since the SD in the present problem is 30, we substitute it for σ to find that $\pm.67 \times 30 = \pm20.10$. The middle 50% of the distribution, therefore, lies between 162 ± 20.10 or between the limits 141.90 and 182.10. (See Fig. 14.)

(b) The middle 75% of the distribution lies 37½% to the right and 37½% to the left of the M. From Table I, we read that a distance of 1.15σ includes the 37½% above the M, and -1.15σ will of course include the 37½% below the M. Substituting $\sigma = 30$, we have that $\pm1.15 \times 30 = \pm34.50$. Accordingly, the middle 75% of the distribution of scores falls between 127.50 and 196.50 (162 ± 34.50). (See Fig. 14.)

(c) Since we want the percentage of scores that falls *below* the score of 112, we must go down to 111.5, the lower limit of score 112. This takes us -50.5 $(111.50 - 162)$ below the M, and this deviation divided by 30 gives -1.68. From Table I we find that 45.35% of the distribution falls between M

* Actually $.6745\sigma$; see p. 77.

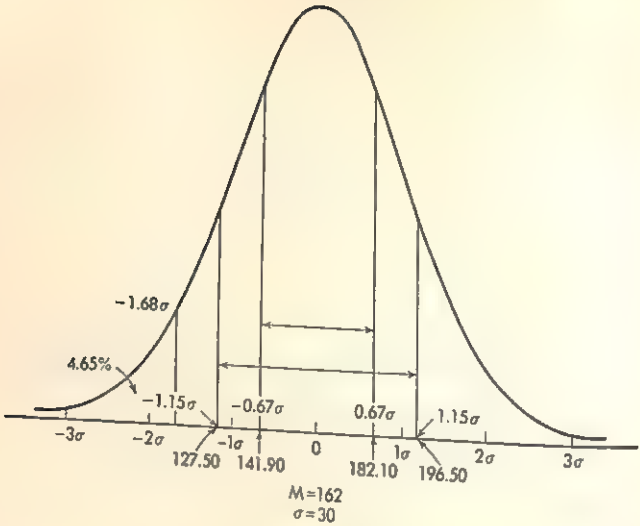


Figure 14

and -1.68σ and accordingly 4.65% of the distribution must fall to the left of this point, or below score 112.

(2) TO SEPARATE A LARGE GROUP INTO SUBGROUPS IN TERMS OF SOME MEASURED TRAIT

Example (1): Grades of A, B, C, D and F are assigned to a class of 500 freshmen enrolled in English I. If achievement in English can be assumed to be normally distributed, how many freshmen should receive each grade?

On the assumption of a normal distribution of achievement in English, we may represent each of our 5 grade groups within a normal curve as shown in Figure 15. The base line of the curve may be taken to extend from -3σ to 3σ, i.e., to cover a range of 6σ. If we divide this range by 5, the number of grade groups, 1.20σ is the base line extent allotted to each

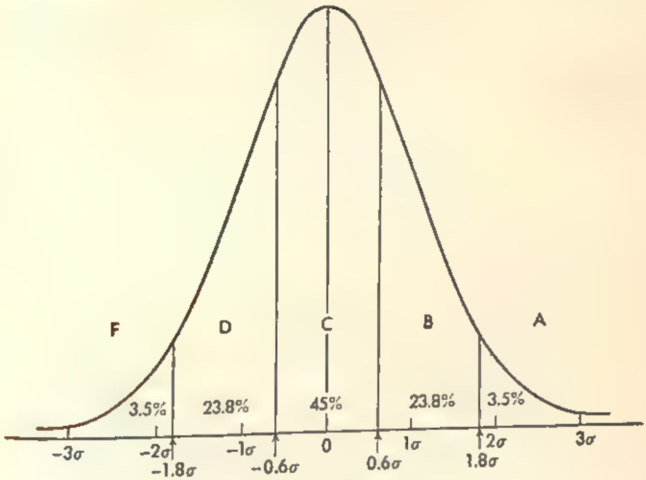


Figure 15

subgroup. The position of each group is shown in Figure 15. The middle or C group covers the area extending $-.6\sigma$ and $.6\sigma$ from the M. Grade group B occupies the area between $.6\sigma$ and 1.8σ; and grade group A, the area between 1.8σ and 3.0σ. The D and F groups occupy positions on the left half of the curve which correspond exactly to those occupied by groups B and A on the right.

To find what percentage of the entire group (N) belongs in A we must determine the per cent of area between 1.8σ and 3.0σ. From Table I we find that 49.87% of N lies between the M and 3.0σ; and 46.41% between the M and 1.8σ. (Note that in both cases we must read area percentages from the M; we cannot subtract the σ-distances and go directly to Table I.) This shows that 3.5% (49.87 − 46.41) falls in that part of the normal curve between 1.8σ and 3.0σ.

For subgroup B, the per cent of the curve lying between 1.8σ and $.6\sigma$ is 23.8% and for subgroup C, the per cent lying

between $.6\sigma$ and $-.6\sigma$ is 22.57×2 or 45.14%. As said before, D and F correspond exactly to B and A. See table below:

A	49.87% (3σ)	− 46.41% (1.8σ)	= 3.5%
B	46.41% (1.8σ)	− 22.57% ($.6\sigma$)	= 23.8%
C	22.57% ($.6\sigma$)	+ 22.57% ($-.6\sigma$)	= 45.1%
D	46.41% (-1.8σ)	− 22.57% ($-.6\sigma$)	= 23.8%
F	49.87% (-3σ)	− 46.41% (-1.8σ)	= 3.5%
			99.7%

Since there are 500 freshmen in the whole class, the numbers in each grade category may easily be found by taking the appropriate percentages of 500:

	A	*B*	*C*	*D*	*F*
Per cent in each grade group	3.5	23.8	45.1	23.8	3.5
Number of students in each grade group *	18	119	226	119	18

Perhaps the point should be stressed that the numbers found in the different subgroups are not *necessarily* the numbers of freshmen who will actually receive the grades of A, B, C, etc. The normal curve provides a mathematical expectation which holds strictly when ability is normally distributed. Various factors will often cause the actual grades to differ more or less from this model. When the assumption of normality is tenable, however, the numbers of freshmen in each of the grade groups (shown in the table above) provide a legitimate expectation or table of *norms*. Wide deviations of grades from expectation constitute a challenge, and would have to be accounted for in terms of selection, teaching and marking standards and the like.

The use of letter grades and ratings is common in industry where workers, supervisors and even executives are classified into efficiency groups in terms of ability, demonstrated skill,

* The end groups (A and F) are adjusted slightly to make the total equal 500.

training, experience. If the trait for which ratings are made can be assumed to be normally distributed, subgroups may readily be set up in which the *range* of talent (σ-distance along the base line) is the same in each group.

NON-NORMAL DISTRIBUTIONS

When either high or low scores occur more frequently than scores in the middle of the score-scale, the frequency distribution will be off center or *skewed*. The extent of skewness may be so slight that the frequency distribution is virtually "normal" and may profitably be treated as normal. But it may happen that asymmetry in the distribution is so pronounced that we cannot legitimately assume normality. Figures 16 and 17 picture frequency distributions which are skewed *negatively*, to the left, and *positively*, to the right. Note that when

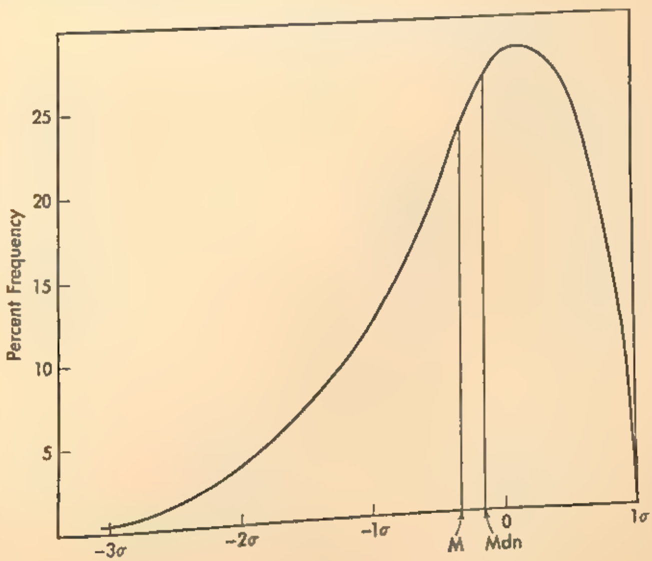


Figure 16

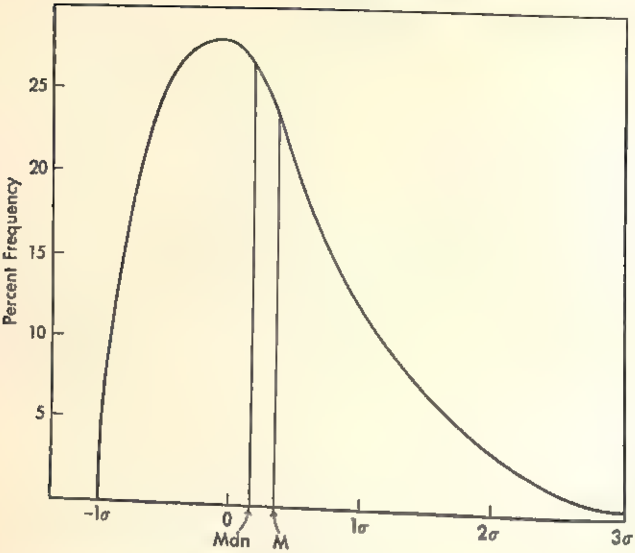


Figure 17

skewness is negative, the *M* lies to the *left* of the *Mdn*: when skewness is positive, the *M* lies to the *right* of the *Mdn*. The degree of skewness in a distribution can be computed mathematically; and in more advanced books methods are given whereby we can determine whether skewness is so large as to preclude the assumption of normality.

The psychologist strives to build his test so that it will yield a normal distribution of scores. This is done on the reasonable assumption that a good test will place most candidates along the middle of the scale, and relatively few at either the high or the low ends. If most of those who take a test score very high or very low—if the distribution is badly skewed—it is obvious that the test is either too easy or too hard for the group as a whole and should be revised so as to give a better, i.e., a more nearly normal, distribution. The test-maker does not assume forthwith that the symmetry of his test distribu-

tion *proves* underlying normality in the trait measured by the test. Instead, he takes the normality in the distribution of test scores to be an index of his success in building a suitable test. We do not know how mental abilities or personality traits are distributed in the general population, but only how our measures of them are distributed. In many cases, the assumption of normality in the distribution is reasonable, and the normal curve provides a useful model. It is this utility of the normal distribution which justifies its wide use.

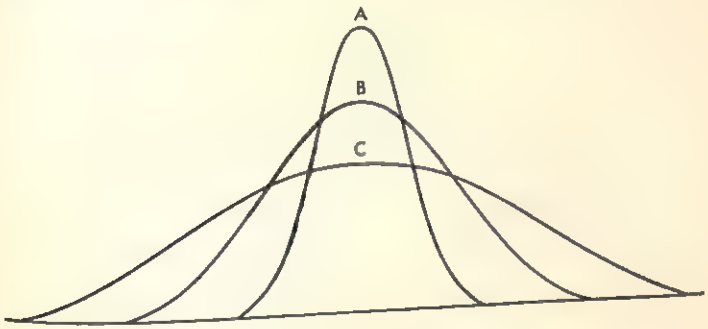


Figure 18

Distributions are not only skewed or off center; in terms of the normal curve model, they may also be too peaked or too flat. The term *kurtosis* refers to this characteristic. Figure 18 shows three distributions: *A* is more peaked than the normal and is called *leptokurtic; C* is flatter than the normal and is called *platykurtic*. The middle distribution, *B*, is *mesokurtic* or normal. It is possible to measure mathematically the degree of kurtosis (with respect to the normal curve) exhibited by a frequency polygon or histogram. And it is sometimes worth while and important to do so.

7.

TESTING EXPERIMENTAL HYPOTHESES

An experiment is set up to test an hypothesis, or to answer a question that arises from the hypothesis. Hypotheses may grow out of hunches and sudden insights. Most often, however, an hypothesis has its origin in some theory and is based upon long-continued observation and experience; or it evolves from other experiments in the same field. An hypothesis is usually stated as a general proposition out of which arise quite specific questions. For example, we may want to know whether method A of teaching reading in Grade 1 is better than method B; whether playgrounds and recreation centers significantly reduce juvenile delinquency in cities over 250,000; whether projective tests predict the outcome of a system of mental therapy better than performance tests; how widely different culture patterns affect personality traits in two primitive societies; the relative importance of psychological factors in distance and depth perception; and whether vocational tests are valuable in forecasting success as a salesman. Queries like these are attacked through the experimen-

tal method, and often require fairly elaborate experimental designs.

In testing an hypothesis, one or several "experimental factors" or independent variables * are tried out at various strengths in the different experimental groups; and the net effect of these factors upon behavior is checked against their absence in a "control" group. It is necessary, therefore, for the experimenter to evaluate the differences found. A difference in the performance of two groups is judged to be "significant" or stable when it is *real*: when it is too large to be attributed solely to "chance" and hence can be expected to stand up or be maintained in subsequent trials. A non-significant difference as between two groups is one which can be dismissed as being too small to provide an appreciable distinction between the groups.

Various sorts of differences between two groups may be of interest to an experimenter. But the difference most often examined is that between two means, and in many respects this is the most important difference. Accordingly, the following sections will be concerned mainly with (1) the conditions under which the difference between two means can be tested legitimately; and with (2) the contingent question of when such differences can be regarded as significant, and when they may be regarded as of no consequence. Before we can test the difference between two means, we must know how to measure the *stability* of each of the means which enter into the difference. This requires a brief look at sampling theory.

* An independent variable is a method, a technique, a set of conditions, or a requirement of some sort which conceivably can influence behavior. Independent variables are under the control of the experimenter and may be applied or withheld at will. Scores or other measures which reflect the influence or strength of the independent variables are called "dependent variables."

Sample and Population

A sample is a group drawn from a larger entity called a population.* In order to infer from the performance of a sample (its M, for instance) what performance can be expected from the population, the sample must be *representative* of its population. To illustrate what is meant by "representative sample and population" suppose that we have administered a standard test of arithmetic to 150 pupils and have computed the M and σ. These school children, let us say, are sixth graders drawn from the several elementary schools in a city of medium size. Now, provided our sample of 150 is a good cross-section of *all* of the sixth graders in the city, we can take its mean to be characteristic of the typical performance of sixth graders in this city. Our sample clearly will *not* be representative † of the population of sixth graders if we have tested only bright children or only dull children, or if our group is drawn mainly from poor neighborhoods, or from schools which have superior teachers. It can be shown mathematically that the best way to obtain a representative sample is to draw its members at *random* from the population. Our first problem, therefore, is to make sure that we have a random sample.

A random sample is one in which (1) every person in the population has the same chance of being drawn or of being included, and in which (2) no single choice forces or deter-

* The members of a population are always alike in some significant respect: all ten-year-old school children, all registered Democrats, all freshmen in midwestern colleges.

† Unrepresentative groups are called biased samples. Sometimes polls go wrong because of biased sampling. If all the voters questioned before an election are urban dwellers, or wealthy people, or socialists, they will not be representative of voters in general. The *Literary Digest* poll of 1936 which predicted the election of Landon over Roosevelt is perhaps the best-known instance of a badly biased sample. This was true despite the fact that the sample was large—more than 2,000,000.

mines another choice (this might happen if taking Mary required that we also take her sister, Margaret). While samples drawn at random may—and usually will—differ slightly, in the long run random selection is our best assurance of a true cross-section of the population. Various devices, not all of which apply in every situation, have been employed to guarantee a random sample. In the problem stated above, for example, the experimenter would try to select children proportionally from all of the elementary schools in the city, thus including all intellectual and socioeconomic levels. When the population is on file (telephone directory or civil service list, for instance) every twentieth or even every five hundredth name might be chosen depending upon the size of the sample wanted; or a table of random numbers * might be used. Once an adequate sample has been assembled, the degree to which its mean represents the population mean can be estimated from the standard error of the mean, designated σ_M or SE_M.

The Standard Error of the Mean (SE_M)

When the sample is random, the standard error of its mean may be found from the following formula:

$$\sigma_M \text{ or } SE_M = \frac{s}{\sqrt{N}} \qquad (15)$$

(standard error of a mean)

in which

s = the SD computed by formula

$$s = \sqrt{\frac{\Sigma x^2}{(N-1)}} \text{ instead of } \sqrt{\frac{\Sigma x^2}{N}}$$

N = size of the sample

* An elementary treatment of the use of random numbers will be found in Helen M. Walker, *Elementary Statistical Method* (New York: Henry Holt and Co., 1943).

The student should mark carefully the use of s in the formula for the SE_M instead of the usual SD. It may be shown mathematically that $(N - 1)$ used in the denominator of the equation for the SD gives a better estimate of the SD in the population than does N; and this correction is especially important when N is small. When N is less than 30, use of $(N - 1)$ makes a considerable difference in the size of the SD, and s should always be used. For large N's the correction effected by using $(N - 1)$ instead of N is negligible.

We may illustrate the application of formula (15), by assuming that the statistics on the arithmetic test given to our 150 sixth graders (see p. 90) are as follows: (N is, of course, known)

$$\text{Mean} = 82$$
$$s = 20$$
$$N = 150$$

Substituting in (15), $SE_M = 20/\sqrt{150}$ or $20/12.25 = 1.63$. This SE_M is to be interpreted in the following way. The frequency distribution of means obtained from random samples drawn from the same parent population can be represented by a normal curve with the "true" or population mean at the center.* Figure 19 shows such a sampling distribution in which the true or population mean is placed at the middle of the curve, and the standard deviation is σ_M or SE_M. The means of the various samples fall around the true mean, more of them being close to than far away from this central point. The heights of the perpendiculars from the base line to the curve give the frequency of sample means at the different points. The probable divergence of our obtained mean of 82 from the population mean can be expressed in probability terms by taking account of the area under the normal curve.

* This is strictly true when N is larger than 30, and approximately true for smaller N's.

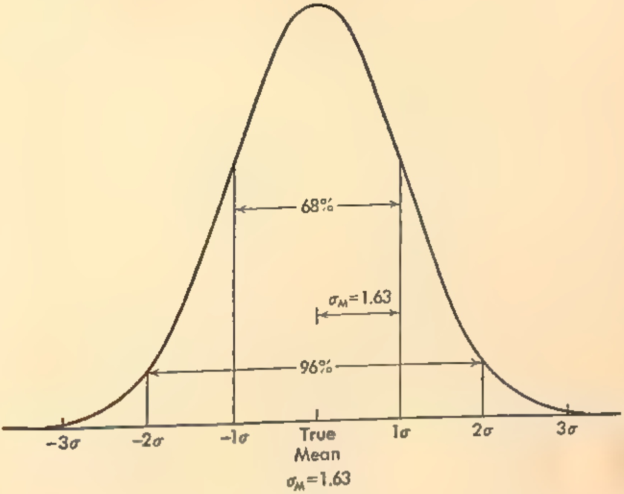


Figure 19

We know from Table I that 68% of the area under the normal curve lies within $\pm 1\sigma$ from the mean. Hence we may say that the chances are 68 in 100 or roughly 2 in 3 that our sample mean of 82 does not miss the population mean by more than ± 1.63 (i.e., by more than $\pm 1\sigma_M$). † Moreover, the chances are 96 in 100 that our mean of 82 does not miss the population mean by more than ± 2 SE_M or by more than ± 3.26 ($\pm 1.63 \times 2$: see Table I). It thus appears that the *stability* of an obtained mean—its divergence from the population mean—can be expressed in probability terms; and that stability varies *directly* with the size of the sample (N) and *inversely* with the size of $s(SD)$. The smaller the SD (less the variability of scores) and the larger the N the smaller the SE_M. The SE_M has a meaning and application in its own right. But its chief value for present purposes lies in the fact that it constitutes

† The normal curve as used heretofore has represented a frequency distribution of scores with the M as the central point. In Figure 19 the M is the "true" or population mean, and the "scores" are sample means. In such a frequency distribution, the $SD = SE_M$.

a necessary intermediate step in the process of calculating the significance of the difference between two means.

The Significance of the Difference between Two Means

How to ascertain the degree of confidence which we can put in the difference obtained between two means can best be shown by an example.

Example (1): In a course in experimental psychology, the members of a class of 20 students were assigned at random to two groups—an experimental and a control. Both groups then undertook a mirror drawing experiment (tracing a series of mazes in a mirror). The experimental group was given specific instructions on how to avoid errors, whereas the control group was allowed to proceed "on its own." Errors in the two groups were as shown below; two members of the control group did not complete the experiment and their records had to be discarded, thus reducing the number in this group to 8.

Group 1 (experimental) $N_1 = 10$

Scores *(X)* *(errors)*	x_1	x_1^2
18	2	4
17	1	1
16	0	0
12	−4	16
19	3	9
15	−1	1
19	3	9
14	−2	4
18	2	4
12	−4	16
10 | 160		64
$M_1 = 16$		

Group 2 (control) $N_2 = 8$

Scores *(X)* *(errors)*	x_2	x_2^2
15	−3	9
18	0	0
19	1	1
18	0	0
19	1	1
20	2	4
19	1	1
16	−2	4
8 | 144		20
$M_2 = 18$		

$$N_1 - 1 = 9$$
$$N_2 - 1 = \underline{7}$$
$$16 \text{ degrees of freedom } (df)$$

$$s = \sqrt{\frac{64 + 20}{9 + 7}} = \sqrt{\frac{84}{16}} = 2.29 \qquad \text{by (16)}$$

D (difference between the two means) $= 2.0$

$$SE_D = \sqrt{\frac{(2.29)^2}{10} + \frac{(2.29)^2}{8}} = 1.09 \qquad \text{by (17)}$$

CR or $t = 2.0/1.09 = 1.83$

When groups are small—as here—instead of computing the SD for each set of scores (X_1 and X_2) separately, we can get a better estimate of the population SD by pooling the sums of squares around the means of the two groups and dividing this total by the numbers of degrees of freedom $[(N_1 - 1) + (N_2 - 1)]$. One degree of freedom (df) is lost in each group, and hence 1 is subtracted from each N.* The formula for s (best estimate of the SD) when sums of squares are pooled is

$$s = \sqrt{\frac{\Sigma x_1^2 + \Sigma x_2^2}{(N_1 - 1) + (N_2 - 1)}} \qquad (16)$$

(SD obtained by pooling the sums of squares in two groups)

Substituting in the formula, $s = 2.29$. The standard error of M_1 (the experimental group) is $s/\sqrt{N_1}$ or $2.29/\sqrt{10}$ by formula (15). And the standard error of M_2 (the control group) is $s/\sqrt{N_2}$ or $2.29/\sqrt{8}$ by formula (15). From these SE's of our two means, we can compute the SE of the difference (D) between the two means by the formula:

* Degrees of freedom is a mathematical concept which is treated at length in advanced texts. For present purposes, it is sufficient to know that the term *df* refers to the number of restrictions placed on the data; or the freedom left when restrictions are imposed.

$$\sigma_D = SE_D = \sqrt{SE^2_{M_1} + SE^2_{M_2}} \tag{17}$$

(SE of a difference when two means are independent)

in which $SE_{M_1} = s/\sqrt{N_1}$ and $SE_{M_2} = s/\sqrt{N_2}$

In Example (1), SE_D is 1.09; and we are now ready to test the significance of the difference between the two means, namely, (18 − 16) or 2. First, we must compute a "critical ratio" or t-ratio defined as follows:

$$CR \text{ or } t = D/SE_D \tag{18}$$

(critical ratio for testing the significance of the difference between two independent means)

Substituting $D = 2$ and $SE_D = 1.09$, we get a t of 1.83. To evaluate the significance of this t we must go to Table II in the Appendix with the given t and the df in our problem. The degrees of freedom are 16 and the critical ratio or t is 1.83. For $df = 16$, we read in the .05 column of Table II that t is 2.12; and in the .01 column that t is 2.92. Our computed t of 1.83 does not reach the .05 level of 2.12, much less the .01 level of 2.92. Hence, we may conclude that the obtained mean difference of 2 found in Example (1) is *not significant*, even at the .05 level. There is no evidence that the two groups really differ in mirror drawing and that the specific instructions given the experimental group had any appreciable effect upon its performance.

Meaning of the Critical Ratio (t) and the Standard Error of a Difference (SE_D)

In order fully to understand why the conclusion "a non-significant difference" was reached in Example (1) above, it will be necessary for us to consider (1) the meaning of the term *null hypothesis* and (2) to digress again into certain

aspects of sampling theory. In any experiment, the experimental group is *expected* to perform differently from the control—to react faster (or slower), learn more (or less) efficiently, reveal personality or attitude differences. Moreover, in comparing any two groups—boys and girls, for example, on a mechanical aptitude test—one group is expected to perform better than the other. The null hypothesis in its most common form asserts flatly that the true mean difference between the two groups being compared is zero; and that the obtained difference (if one has been found) is inconsequential and could well be zero. The purpose of an experiment is to give the facts a chance to disprove or confirm (fail to disprove) this null hypothesis. In rejecting a null hypothesis, we assert that the difference obtained *is* significant, that it indicates the existence of a true difference greater than zero. In accepting the null hypothesis, on the other hand, we concede that there is no reason to suspect—as far as our data are concerned—that the true difference is *not* zero.

In Example (1) the obtained difference of 2.0 was evaluated against the null hypothesis, namely, that the true difference is zero. The test of the null hypothesis is represented graphically by the curve in Figure 20. This sampling distribution shows the spread of obtained differences around the hypothetical difference (D) of zero, that is, it shows the way in which the obtained D's might be expected to occur under the null hypothesis.

Table II lists, for various degrees of freedom, the values of t which correspond respectively to the .05 and the .01 levels of significance. The first point cuts off 2½% from each end of the curve, the second cuts off ½ of 1% from each end. Entering Table II with the 16 degrees of freedom in our problem, we read that the .05 value is 2.12 and the .01 value is 2.92. The first entry means that 5% of t-ratios (obtained from repeating experiments like ours) can be expected to *exceed* 2.12 even

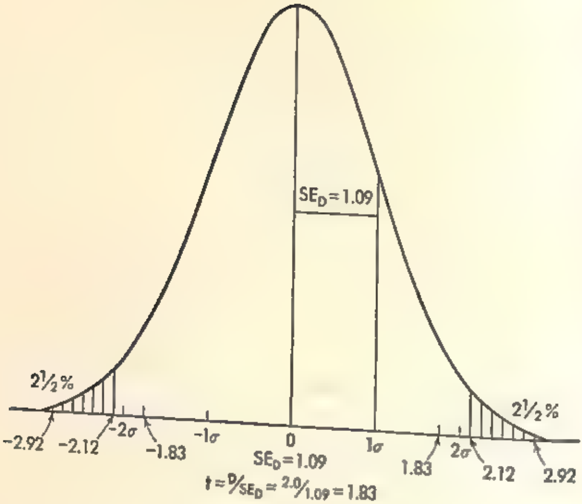


Figure 20

when the null hypothesis is true; and the second entry means that 1% of such t-ratios can be expected to exceed 2.92 when the null hypothesis is true. Our obtained difference of 2.0—measured in terms of its SE_M and expressed as a t-ratio of 1.83—falls short of 2.12. Hence our t does not reach the .05 point and must be marked "not significant at the .05 level." Had our t exceeded the 5% point, we would have concluded that a difference as large as 2.0 occurs too *infrequently* "by chance" to be written off as due to random errors. On the other hand, when our obtained t falls short of the 5% point (as it does here), we regard the obtained difference as representing merely a chance deviation from the true difference of zero.

When we accept the 5% level of significance, we are saying in effect that we are willing to gamble on being wrong *once* in twenty trials—or five times in a hundred trials. And when we accept the 1% level of significance, we are taking the risk of being wrong only once in a hundred trials. Sometimes a t falls

between the two levels—the 5% and the 1%—in which case our difference is said to be significant at the .05 but not at the .01 level. Only these two levels of significance are ordinarily needed. Both are, to be sure, arbitrary, but they provide reasonable standards for experimental work. The .05 level permits a considerable degree of confidence in a difference; the .01 level a much higher degree of confidence.

A second example will make clearer the procedure to be followed in accepting or rejecting the null hypothesis.

Example (2): The following scores were achieved on a Vocabulary Test by men and women students in an arts college. Is the difference in mean score significant?

	Male	*Female*	*Difference*
Mean	31.97	33.39	1.42
SD	5.50	5.20	
N	713	287	

$$SE_D = \sqrt{\frac{(5.50)^2}{713} + \frac{(5.20)^2}{287}}$$

$$= .370$$

$$CR \text{ or } t = \frac{1.42}{.37} = 3.84$$

$$df = 712 + 286 = 998$$

These samples are so large that it would make no appreciable difference whether we used s or SD. $SE_{M_1} = 5.50/\sqrt{713}$ and $SE_{M_2} = 5.20/\sqrt{287}$; thus we have from formula (17) that $SE_D = .370$. The critical ratio or t is 1.42/ .370 or 3.84, and the degrees of freedom (df) are 712 + 286 or 998 $[(N_1 - 1) + (N_2 - 1)]$. From the last line in Table II, we find the .05 value to be 1.96 and the .01 to be 2.58. Since our t of 3.84 exceeds .01 by a considerable margin, we mark our difference "highly significant" or significant *beyond* the .01 level. We may now reject the null hypothesis with great con-

fidence, even though the obtained difference of 1.42 is actually quite small. The entries in the bottom line of Table II give the .05 and .01 values for very large N's. For practical purposes we may use these two points for *any* samples greater than 100 (or when the *df* are larger than 100) with little margin of error. If the student will check the two points ± 1.96 and ± 2.58 in Table I, he will find that the first marks off 5% in the two extremes (tails) of the normal curve (2½% in each end), and that the second point marks off 1% in the two extremes of the normal curve (½ of 1% in each tail). For samples larger than 30, the sampling distributions of differences (under the null hypothesis) are essentially normal curves (see p. 92).

The Significance of the Difference between Means Obtained from the Same Group upon Two Occasions

In the two samples given in previous sections, the means being compared were obtained from different groups and hence were *independent* or uncorrelated. In many experiments, however, when the *same* group has been measured upon two or more separate occasions, the means are not independent, one mean being very probably correlated with the other. School children, for example, are given educational achievement tests in the fall and again in the midyear; or subjects in a learning experiment are given repeated trials at the same task. In cases like these, there is always the possibility of correlation (sometimes high) between scores achieved by the group upon the separate trials, and the means are usually not independent.

When a later performance is compared with some earlier trial, the group is said to furnish its own control: that is, it is both the experimental and the control group. One useful method of comparing means under these conditions is to find,

first, the difference in score for each person. We may then test the significance of the mean difference against the null hypothesis using the method of the last section. An example will serve to illustrate this "difference method."

Example (1): At the beginning of the term, a class of 10 students in a course in mental hygiene was asked to fill out an "adjustment inventory" covering 200 behavior items. At the end of the semester, the same 10 students were again asked to fill out the inventory, and the answers on the two occasions were compared. Is there any evidence of improvement in adjustment?

Students	*1st Administration of Inventory*	*2nd Administration of Inventory*	$D(X)$	x	x^2
A	42	36	6	2.6	6.76
B	33	30	3	—.4	.16
C	38	31	7	3.6	12.96
D	46	45	1	—2.4	5.76
E	50	42	8	4.6	21.16
F	34	36	—2	—5.4	29.16
G	47	40	7	3.6	12.96
H	54	56	—2	—5.4	29.16
I	31	30	1	—2.4	5.76
J	36	31	5	1.6	2.56
			10 $\overline{)34}$		126.40

$$\text{Mean}_D = 3.4$$

$$s_D = \sqrt{\frac{126.40}{(10\text{-}1)}} = \sqrt{14.044} = 3.75$$

$$SE_{M_D} = \frac{3.75}{\sqrt{10}} = 1.19$$

$$CR \text{ or } t = \frac{(3.4 - 0)}{1.19}$$

$$t = 2.86$$

$$df = (N - 1) = 9$$

The score made by each student at the end of the course is subtracted from his initial score giving a column of 10 differences. Eight of the students improved—made *lower* scores, thus showing better adjustment—on the second administration of the inventory; and two had higher scores. The mean of the 10 differences is 3.4 and the s around this mean is 3.75 $\left(\sqrt{\frac{\Sigma x^2}{(N-1)}}\right)$. The standard error of the mean difference or $SE_{M_D} = 1.19$ by formula (15). In order to test the obtained mean difference against the null hypothesis (that is, against the hypothesis that the true mean difference is 0), we compute a critical ratio or t which equals 2.86. Note that the deviation of the obtained M of 3.4 from the expected M of 0 furnishes the D, which is evaluated in terms of SE_{M_D}. Figure 21 pictures the situation graphically.

For (10-1) or 9 degrees of freedom, the .05 and the .01 points are 2.26 and 3.25, respectively (see Table II). As shown in Figure 21, our obtained t of 2.86 exceeds 2.26, but

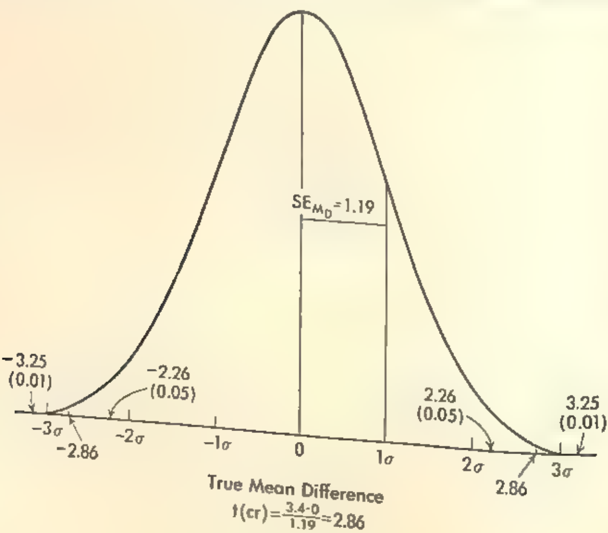


Figure 21

does not reach 3.25. Hence, our mean difference of 3.4 must be marked "significant at the .05 but not at the .01 level." On the present evidence, we may be fairly confident that students will improve in adjustment score following a course in mental hygiene.

The Significance of the Difference between Two Percentages

In a considerable number of experimental problems, we are able to compute the percentages in two or more groups that exhibit a certain behavior, when it is not feasible to measure the behavior itself in terms of test scores. This is especially true in social behavior, which is often all-or-none or present or absent. For example, the incidence of smiling in infants at different age levels, of aggressiveness in preschool children, of snobbery in teen-agers—such behavior can be most readily recorded by counting the numbers in the various groups that reveal it. The significance of the difference between the percentages of two groups who exhibit a given behavior may be tested against the null hypothesis—the hypothesis that no real difference exists between the two groups. The formula for the standard error of a percentage difference is

$$SE_{D_{\%}} = \sqrt{PQ\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \tag{19}$$

(*SE* of the difference between two independent or uncorrelated percentages)

in which

P = mean of the percentages in the two groups exhibiting the behavior

$Q = (1 - P)$

N_1 = number of cases in Group 1

N_2 = number of cases in Group 2

The pooled estimate of P is found by the formula:

$$P = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}$$

and $$Q = (1 - P)$$

Two examples will illustrate the use and interpretation of formula (19).

Example (1): In a poll taken in City A, 46% of a sample (presumably random) of 200 registered Democrats were recorded as favoring a certain issue. In the same city, 52% of a sample of 240 registered Republicans also favored the issue.

May we conclude that Democrats and Republicans differ in the strength of their attitude toward this policy?

First, we must obtain P—the pooled estimate of the percentage in the two groups who on the average favor the issue—by the formula:

$$P = \frac{200 \times .46 + 240 \times .52}{(200 + 240)}$$
$$= 49.3\%$$
$$Q = (1 - P) = 50.7\%$$

By formula (19), we have that

$$SE_{D_{\%}} = \sqrt{49.3\% \times 50.7\% \left[\frac{1}{200} + \frac{1}{240}\right]}$$
$$= 4.8\%$$

D (the difference between the two per cents) is
$$52\% - 46\% \text{ or } 6\%$$

$$t = 6\% / 4.8\% = 1.25$$

The df is so large, namely, 438 [(200 − 1) + (240 − 1)], that we may safely use the last line of Table II in determining

the significance of our difference of 6%. The t at the .05 level is 1.96, and at the .01 level is 2.58. Our critical ratio of 1.25 fails to reach even the smaller of these two values and hence must be marked not significant. Accordingly, we accept the null hypothesis and conclude that on the present evidence there is no real difference between Democrats and Republicans in the extent to which they favor the given issue.

Example (2): Suppose that 20 students who have come down with common colds are treated with vaccine A and that 80% are well within two weeks. At the same time, 20 other students, matched for age and sex with the first 20, come down with colds and are left untreated. Of this group, 70% are well within two weeks. Is there any evidence that vaccine A is really effective in curing the common cold?

First, we find P:

$$P = \frac{20 \times 80\% + 20 \times 70\%}{40}$$

$$= 75\%$$

$$Q = (1 - 75\%) = 25\%$$

$$SE_{D\%} = \sqrt{75\% \times 25\% \left[\frac{1}{20} + \frac{1}{20}\right]} \qquad \text{by (19)}$$

$$= 13.7\%$$

D (the difference between the two per cents) is
$(80\% - 70\%)$ or 10%

and the critical ratio is

$$t = 10\%/13.7\% = .73$$

For 38 degrees of freedom (Table II) the .05 and .01 points are 2.02 and 2.71, by interpolation.

It is evident that our t of .73 fails to reach 2.02 and that the obtained difference is far from being significant, despite its size (10%).

8.

CORRELATION

The term correlation refers to the degree of correspondence or relationship between two sets of test scores or other measures. Degree of correspondence is expressed by the coefficient of correlation (r) along a scale which extends from 1.00 to −1.00. A coefficient of 1.00 denotes perfect relationship: a theoretical upper limit approached but rarely reached with real data. An $r = .00$ implies no true relation, whereas an r of −1.00 indicates perfect but *inverse* relationship. Between 1.00 through .00 to −1.00 different degrees of correspondence are expressed by such coefficients (decimals) as .60, −.30, .20, etc.

We may illustrate what is meant by a *positive* relationship by supposing that 10 high-school boys have been given two tests, one in algebra and the other in natural science. If the correspondence of the two sets of scores is perfect or 1 to 1 (highest boy in algebra the highest in science, next highest boy in algebra the next highest in science and so on) the r is 1.00. If 2 or 3 boys fail to show complete correspondence in score position (as is likely), the r will still be high (.80, perhaps) but no longer perfect.

When standing on one test is not reflected at all in

the second test, the $r = .00$. For example, if good basketball players are as often high in their studies as they are low or intermediate, the r between school grade and prowess in basketball will be close to zero. The correspondence between scholarship and number of extracurricular activities is often *negative*: this may mean that the greater the outside activity the student engages in the less time he has for study.

There are a number of methods for computing a measure of correlation. Which to use will depend in part upon the character of the data, whether expressed in scores, categories or ranks. In the present chapter two helpful and widely used correlational methods will be outlined: the *rank-difference* and the *linear* or *product-moment*. Other methods of correlation will be found in more advanced texts.

Computing Correlation from Rank Orders (Rank-Difference Method)

The rank-difference coefficient of correlation is a measure of relationship between the rank orders held by the members of a group in two activities. Differences among people are often expressed by ranking them in 1-2-3 order when the behavior in which we are interested cannot be measured directly. Preschool children, for example, may be ranked for aggressiveness or social adjustment; workers, for industry, initiative and personality traits. Furthermore, objects may be put in order for some attribute. Thus advertisements, pictures, tonal combinations, even automobiles may be ranked in order of merit for buying appeal, esthetic qualities, economy, comfort, etc., when it is difficult or impossible to measure such qualities along some scale. From the differences between the two sets of ranked data, a coefficient of correlation called ρ (read *rho*) can be computed. Rho is a close approximation to r, the "standard" index of relationship between two traits

measured in terms of scores. The example given below will illustrate the computation of rho.

Example (1): A group of 10 seventh-grade children who had earned the scores shown below on a history test were ranked by their teacher for study habits (1 being best; 10, worst). Rank these children in order for achievement in history and compute a rank-order correlation between the two sets of ranks.

(1)	(2)	(3)	(4)	(5)	(6)
Children	*Scores in history*	*Ranks in history*	*Ranks in study habits*	*Diffs. in rank (D)*	*Diffs. in rank squared (D^2)*
John	81	4	3	1	1
William	93	1	1	0	0
Sue	84	2	5	3	9
Ann	65	8.5	10	1.5	2.25
Vince	70	7	7	0	0
Gus	81	4	2	2	4
Mary	65	8.5	4	4.5	20.25
Henry	60	10	9	1	1.00
Fred	73	6	8	2	4.00
Betty	81	4	6	2	4.00
					45.50

$$\rho\ (rho) = 1 - \frac{6 \times \Sigma D^2}{N(N^2-1)} = 1 - \frac{6 \times 45.50}{10 \times 99} = .72$$

The children's history scores, shown in column 2, have been ranked in order of merit in column 3. Column 4 shows the ranks for goodness-of-study habits. Note in column 3 that 3 children—John, Gus and Betty—have the same history score (namely, 81) and that each is given the rank of 4. Since we do not know in what order these 3 children should be put (3, 4, 5) we simply rank all three 4 and Fred, who follows, 6. Again,

Ann and Mary both have the same score (65). Instead of ranking Ann 8 and Mary 9, each girl is ranked 8.5 and the next pupil, Henry, is ranked 10.

The differences (D) between each pair of ranks, without regard to sign, are shown in column 5, and the differences squared (D^2), in column 6. The sum of these squared rank-differences (D^2) together with N, the size of the sample, are put in the rank-difference formula to give the coefficient ρ:

$$\rho = 1 - \frac{6 \times \Sigma D^2}{N\,(N^2-1)} \tag{20}$$

(rank-difference coefficient of correlation ρ, *rho*)

In the formula ΣD^2 = sum of the squared differences (D)
N = size of the sample

Substituting for $\Sigma D^2 = 45.50$ and for $N = 10$, we get a ρ of .72. This coefficient indicates a substantial correlation between history test score and goodness-of-study habits: a not unexpected finding. Rho is valuable as an exploratory device, and is often used in pilot or preliminary studies as a means of detecting whether there is any correlation present. Rho is also useful when the behavior in which we are interested cannot be measured directly, but it is possible to put individuals in rank order with respect to it. Rho is not recommended when N is greater than 25 or so, as ranking in 1-2-3 order is then a difficult task.

LINEAR CORRELATION

The Correlation Coefficient, r

The correlation coefficient r provides a quantitative measure of the relationship between two variables X and Y (a variable is a set of scores or other measures which vary along

some scale from high to low). The coefficient r is called a measure of "linear" correlation because it describes a relationship which is expressed by a straight line. Figure 22 shows graphically a case of linear relationship. The scores made by each of 9 subjects upon two tests have been plotted against each other: each point, therefore, represents a *pair* of scores. In the diagram, for example, score 15 in Test X is plotted against score 5 in Test Y, score 20 in X against score 10 in Y, and so on to score 60 in X which is plotted against score 25

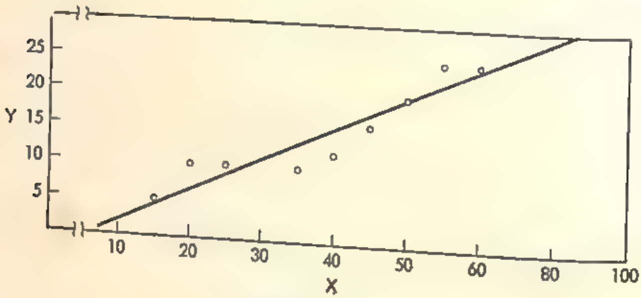


Figure 22

in Y. A straight line has been drawn in "by eye" to indicate the trend of the points. This line is drawn through or as near as possible to the 9 separate points. It shows that, despite fluctuations, the relationship between scores on the two tests can be well described by a straight line. As scores go up in one test, they tend to go up in the other.

Calculation of r Directly from Paired Scores

The formula for r may be written in a number of ways of which the following is one of the most useful:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} \qquad (21)$$

[coefficient of correlation r when deviations (x and y) are taken from the M's of X and Y]

in which

Σxy = sum of the products of deviations x and y
Σx^2 = sum of the squared deviations in X from M_x
Σy^2 = sum of the squared deviations in Y from M_y

The application of formula (21) is illustrated in Example (1).

Example (1): The paired scores in Table 16 represent the achievement of 14 sixth graders upon tests of arithmetic reasoning (X) and arithmetic computation (Y). Compute the correlation by formula (21).

An outline of the calculations in Table 16 is given in the following steps.

Step 1. Find the M's of the two tests, X and Y. The mean of the arithmetic test (M_x) is 52; and the M of the computation test (M_y) is 26.5.

Step 2. Enter the deviations of each score X (arithmetic reasoning) from M_x (52) in column 3, headed x. Also enter the deviations of each score Y (arithmetic computation) from M_y (26.5) in column 4. Note that deviations are minus when scores are below the mean.

Step 3. Square the x's and the y's and enter these squares in columns 5 and 6, headed x^2 and y^2. Total these columns to get Σx^2 and Σy^2.

Step 4. Multiply corresponding x's and y's with *due regard for sign*, and enter the products under xy in column 7. Total the xy to get Σxy.

Step 5. Substitute for Σxy, Σx^2, Σy^2 in formula (21) to obtain $r = .68$ as shown in Table 16.

TABLE 16

To Illustrate the Computation of r from 14 Pairs of Ungrouped Scores, Deviations Being Taken from the Means of the Two Tests

Names	(1) Arithmetic reasoning (X)	(2) Arithmetic computation (Y)	(3) x	(4) y	(5) x^2	(6) y^2	(7) xy
Robert	43	25	−9	−1.5	81	2.25	13.5
Betty	56	34	4	7.5	16	56.25	30.0
Shirley	45	20	−7	−6.5	49	42.25	45.5
Warren	50	20	−2	−6.5	4	42.25	13.0
John	36	25	−16	−1.5	256	2.25	24.0
Georgia	58	30	6	3.5	36	12.25	21.0
June	55	31	3	4.5	9	20.25	13.5
Susan	61	28	9	1.5	81	2.25	13.5
William	46	23	−6	−3.5	36	12.25	21.0
Laurence	64	33	12	6.5	144	42.25	78.0
Jean	46	20	−6	−6.5	36	42.25	39.0
Janet	62	34	10	7.5	100	56.25	75.0
James	50	25	−2	−1.5	4	2.25	3.0
Vincent	56	23	4	−3.5	16	12.25	−14.0
	728	371			868	347.50	376.0
Means	52	26.5			Σx^2	Σy^2	Σxy

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{376.0}{\sqrt{868 \times 347.50}} = .68 \qquad \text{by (21)}$$

$$\sigma_x = \sqrt{\frac{868}{14}} = 7.87 \qquad \text{by (7)}$$

$$\sigma_y = \sqrt{\frac{347.50}{14}} = 4.98 \qquad \text{by (7)}$$

An alternate formula for r is the following:

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y} \tag{22}$$

(r when deviations are taken from the *M's* of the two distributions)

This is the formula most often seen in textbooks on psychology; and it is the one generally used to demonstrate the meaning of correlation. In order to apply it to the data in Table 16 we must compute σ_x (the *SD* of the *X* distribution) and σ_y (the *SD* of the *Y* distribution). By formula (7), page 52, $\sigma_x = \sqrt{868/14}$ or 7.87; and $\sigma_y = \sqrt{347.50/14}$ or 4.98. Substituting in formula (22) we have that

$$\text{r} = \frac{376.0}{14 \times 7.87 \times 4.98} = .68 \text{ (to two decimals)}$$

checking the result from formula (21).

Calculation of r from a Correlation Table

When N is small either formula (21) or (22) will be satisfactory for the computation of r. If N is large, however, both time and computational labor can be saved by computing r from a correlation diagram like that shown in Figure 23.* In cases where N is very large, the use of computing machinery is almost a necessity.

Figure 23 represents by means of a chart or scattergram the paired scores made by 42 eighth grade children upon tests of reading (X) and composition (Y).

The paired scores of each pupil are represented by an entry in the center of one of the cells of the diagram. For example, the two children in the second row from the top and fourth

* An $N = 42$ is not, of course, especially "large." A relatively small group was selected to illustrate the calculations in figure 23 so that the computations would not obscure the method.

column over from the left have reading scores (X) in interval (45-49) and composition scores (Y) in interval (78-83). In the bottom row, the four pupils who scored in interval (54-59) in Y are distributed over 3 intervals (columns) in X:2 in (30-34), 1 in (35-39) and 1 in (40-44). Steps in the computation of r from a correlation table are set out briefly below. Reference is made to Figure 23 throughout.

Step 1. Construct, first, a correlation diagram or two-way table. Enter each pair of scores in its appropriate cell as a tally, and later consolidate all entries into a single figure. This gives us a correlational diagram.

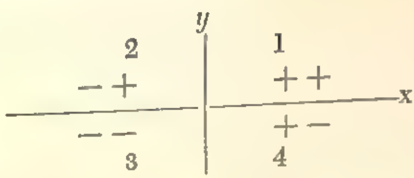
Step 2. Add the columns and rows and sum over the table to find the total frequency in X and in Y. Assume an M for Y (AM_y) and draw in double lines on the diagram to mark off the row containing the AM_y. The interval (66-71) in Y with a midpoint of 68.5 was selected, as it is near the center of the distribution and has the largest f (see p. 81). Take deviations from AM_y in units of i and enter in the column y'. Fill in the columns for fy' and for fy'^2 by multiplying corresponding f's and y''s, and y''s and fy''s. The correction in units of interval or c_y is found from the fy' column to be 15/42 or .357. From the column fy'^2 and from the c_y^2 the SD by formula (9) is found to be $\sigma_y = 6 \times 1.394$, or 8.36.

Step 3. The calculations for X repeat those for Y. The AM_x is taken at 42.0, midpoint of column (40-44); and x' deviations have been entered in interval-units at the bottom of the diagram. Fill in the fx' and the fx'^2 rows. Note again that the prime (′) means that deviations have been taken from the assumed M. The c_x is obtained from the fx' row and σ_x from the fx'^2 row. $c_x = 3/42$ or .071; and $\sigma_x = 5 \times 1.280$ or 6.40.

Step 4. So far we have repeated operations (calculations of σ's) learned earlier. Our first new calculation is that for $x'y'$. These entries are found in the following way. In the top row, the cell in the extreme right-hand corner lies 3 interval-units

or 3 columns to the *right* of the interval containing the AM_x (column with double lines). This cell also lies 3 interval-units or 3 rows upward from the row containing the AM_y (row with double lines). Hence the product-deviation of this cell is 3×3 or 9, and since there is only 1 entry in the cell its deviation ($x'y'$) is also 9. The small (9) in the upper right-hand corner of the cell is the product-deviation of the cell, and the entry 9 in the lower left-hand corner is the $x'y'$ of the single score in the cell. The next cell in the top row to the left has a product-deviation of 2×3 or 6 and the two scores in the cell have an $x'y'$ deviation of 6×2 or 12. This cell is 3 rows upward from the row containing AM_y and 2 columns to the right of the column containing AM_x. The third cell over, has a product-deviation of 1×3 or 3 and an $x'y'$ of 3 since there is only 1 score in the cell. The total $x'y'$ for this row is 24.

The column and row selected to contain the AM's divide the correlation table into four quadrants in the following way:



All $x'y'$ in quadrant 1 are positive (+ +); all $x'y'$'s in quadrant 2 are negative (− +); all $x'y'$'s in quadrant 3 are plus (− −); and all $x'y'$ entries in quadrant 4 are minus (+ −). Entries which lie in the column or row containing the AM's have product-deviations of zero, since *one* deviation (either x' or y') is zero.

Step 5. When all of the cells in the diagram have been assigned final entries—in the lower left-hand corner of each cell—the rows are summed and the sum entered in the $x'y'$ column. In the third row from the top, for example, the sum from right to left is $6 : 3 + 2 + 2 + 0 - 1$. In the fifth row from the

top the sum is 4 : $-1+0+3+2$. Total the $x'y'$ to obtain 56.

Step 6. Now substitute 56 for $\Sigma x'y'$; 42 for N; .071 and .357 for the corrections c_x and c_y; 1.280 and 1.394 for the σ's in *units-of-interval* in the following formula:

$$r = \frac{\frac{\Sigma x'y'}{N} - c_x c_y}{\sigma_x \sigma_y} \tag{23}$$

(coefficient of correlation when deviations are taken from the two assumed means)

As shown in Figure 23, $r = .73$.

Note carefully that the $\Sigma x'y'$ and the corrections (c_x and c_y) are all left in *units-of-interval*; and that this is true also of the two σ's (σ_x and σ_y). [σ's are left in interval units *only* in formula (23)]. Leaving $x'y'$ and c_x and c_y in interval-units facilitates computation and as long as the σ's are also left in interval-units does not affect the final result.

Meaning of a Coefficient of Correlation

The question is often raised of how high a correlation coefficient should be in order to be regarded as "significant." It is difficult to answer this question categorically as the *level* of relationship indicated by r depends upon several factors: (1) the absolute size of the coefficient; (2) the purposes for which r was calculated; (3) how our r compares with r's generally found for the variables or traits studied. For the beginner, who would often like definite answers, the following "rules" will serve as a general guide.

r from .00 to $\pm$ .20	very low or negligible
r from $\pm$ .20 to $\pm$ .40	low; present but slight
r from $\pm$ .40 to $\pm$ .70	substantial or marked
r from $\pm$ .70 to $\pm$ 1.00	high to very high

English Composition (Y)	Reading (X) 30-34	35-39	40-44	45-49	50-54	55-59	f	y'	fy'	fy'^2	$x'y'$
84-89				1 (3) 3	2 (6) 12	1 (9) 9	4	3	12	36	24
78-83			1 (0) 0	2 (2) [illegible]	2 (4) 8		5	2	10	20	12
72-77		1 (−1) −1	3 (0) 0	2 (1) 2	1 (2) 2	1 (3) 3	8	1	8 (30)	8	6
66-71		8 (0) 0	6 (0) 0				14	0			0
60-65	1 (2) 2	3 (1) 3	2 (0) 0	1 (−1) −1			7	−1	−7	7	4
54-59	2 (4) 8	1 (2) 2	1 (0) 0				4	−2	−8 (−15)	16	10
f	3	13	13	6	5	2	42		15	87	56
x'	−2	−1	0	1	2	3					
fx'	−6	−13 (−19)		6	10	6 (22) = 3					
fx'^2	12	13		6	20	18 = 69					

$AM_y = 68.5$

$AM_x = 42.0$

$$\sigma_x = 5\sqrt{\frac{69}{42} - .0050} = 5 \times 1.280 = 6.40$$

$$\sigma_y = 6\sqrt{\frac{87}{42} - .1274} = 6 \times 1.394 = 8.36$$

$$c_y = \frac{15}{42} = .357 \quad c_y^2 = .1274$$

$$c_x = \frac{3}{42} = .071 \quad c_x^2 = .0050$$

$$r = \frac{\frac{56}{42} - .071 \times .357}{1.280 \times 1.394} = .73$$

Figure 23

Coefficient of correlation (r) between Reading (X) and Composition (Y)

PREDICTING ONE VARIABLE FROM ANOTHER

Linear correlation (p. 110) implies relationship that can be expressed by a straight line. In every correlation table there are two lines of relation, called regression lines. The first line enables us, knowing a person's score in test X, to predict or estimate his most likely score in test Y. The second line enables us, knowing a person's score in test Y, to predict his most probable score in test X. Ordinarily we use only the *one* regression line, as we are interested in predicting in only one direction. For example, we may want to estimate a boy's probable achievement (Y) in his freshman year from his intelligence test score (X); or the probable performance of a new salesman from a battery of aptitude tests and ratings; or the probable achievement of an athlete in a track meet from a battery of sensory-motor tests. Prediction in the reverse direction would be meaningless. The equation of the regression line for predicting a score in test Y from a known score in test X is

$$Y_{\text{pred.}} = r\,\frac{\sigma_y}{\sigma_x}\cdot X - r\,\frac{\sigma_y}{\sigma_x}\cdot M_x + M_y \qquad (24)$$

(regression equation for predicting Y-scores from X-scores)

in which

r = the correlation between X and Y
σ_y and σ_x = the *SD*'s of the X and Y scores
X = the known score in test X
M_x and M_y = the means of the X and Y scores
$Y_{\text{pred.}}$ = the predicted score in test Y

We may illustrate the use of formula (24) with the data from Figure 23. Suppose that we know a child's reading score (X) and want to know his most likely score in English com-

position (Y). Substituting for r, σ_x *and* σ_y, M_x and M_y * in formula (24) we have

$$Y_{\text{pred.}} = \frac{.73 \times 8.36\,X}{6.40} - \frac{.73 \times 8.36 \times 42.36}{6.40} + 70.64$$

$$\text{or } Y_{\text{pred.}} = .95\,X + 30.25$$

From this equation we can forecast the most likely score of a pupil in test Y (composition) when we know his score in test X (reading). If William, for example, made a score of 44 in reading (X), his most likely score in composition (Y) is 72:

$$Y_{\text{pred.}} = .95 \times 44 + 30.25$$
$$= 72.05 \text{ or } 72 \text{ (to two digits)}$$

A reading score (X) of 36 forecasts a composition score (Y) of 64; a reading score of 56, a composition score of 83 and so on. Prediction of Y-scores from X-scores is, of course, of no practical value for the 42 pupils in our group: we already know their composition (Y) scores. For persons for whom we have only (X) scores, however, the ability to predict probable performance in (Y) from a *known* regression equation may be very valuable indeed. Suppose, for example, that we have established the correlation between a battery of intelligence and achievement tests (X) and freshman grades (Y) for a large group. We may use the regression equation to forecast the probable performance of new students for whom we have *only* the intelligence-achievement scores. Or, knowing the correlation between aptitude tests and job performance, we can forecast the most likely performance of clerical workers from their scores on a battery of clerical aptitude tests; or of medical students from their scores on a battery of medical aptitude tests. In every case, of course, the corre-

* M_x and M_y are computed from the AM_x and AM_y in Figure 23 by use of the *ci*'s of .355 and 2.142. $M_x = 42.00 + .355 = 42.36$, and $M_y = 68.5 + 2.142 = 70.64$.

lation between X and Y must first be established upon a large and representative group. Otherwise the regression equation is of limited value. It is often possible to forecast probable achievement over long periods into the future from tests taking not much over an hour to administer.

How much better can we predict scores when we know the correlation between X and Y than we can without this knowledge? When r is high, prediction is good; in fact, we can estimate without error when $r = 1.00$. For an $r = .00$, the forecast is no better than it would be if we did not have the correlation. Thus, if we are asked to forecast the height of a ten-year-old boy from his weight—and we do not know the r between height and weight—our best estimate is simply the mean height of ten-year-olds. In the absence of any knowledge of the correlation, our best "guess" is always the M_y no matter what the individual's score in X. Correlation coefficients between .00 and ± 1.00 provide estimates by way of the regression equation which are improvements over "guesses," the degree of improvement depending upon the size of the r and the amount of spread (size of σ's) in the correlated tests.

Determining the Significance of a Coefficient of Correlation

The confidence we can put in an r as representing a high or low relationship between two variables is contingent upon several factors. *First*, it depends upon how large and representative of its population our sample is; and *second*, it depends upon the size of the r itself. Table III is useful in evaluating the significance of an obtained r. To use the table, we need only know the number of "degrees of freedom" (see p. 95) in the correlation table and the value of r. In a correlation problem the df are (N-2), where N is the number of X-Y pairs. We enter Table III, therefore, with (N-2) df.

In Figure 23 the df are 40 (i.e., $42 - 2$); and from Table III

opposite 40 in the *df* column we find *r*'s of .30 and .39 in the .05 and the .01 columns. These figures mean that for 40 *df* an *r* must be at least .30 to be significant at the .05 level, and at least .39 to be significant at the .01 level. Said differently, when the *df* are 40, only five times in 100 trials (5% of the time) would an *r* as large as or larger than .30 occur "by chance" *if* the true or population *r* were zero. And only once in 100 trials would an *r* as large as or larger than .39 arise "by chance" if the population *r* were zero. Our obtained *r* of .73 is so much larger than .39 (the higher of the two "standard" *r*'s) that there is little likelihood that the true correlation between reading and composition is really zero. Hence, we reject the null hypothesis (i.e., the hypothesis of no correlation) with great confidence and mark our correlation coefficient as being highly significant. There is small probability that an *r* of this size would arise simply from sampling fluctuations.

It is customary to label an *r* which exceeds the .05 value as significant at the .05 level of confidence; and an *r* which exceeds the .01 value as significant at the .01 level of confidence.* Sometimes an *r* falls in between these two points. For example, an *r* of .40 for $df = 28$ is significant at the .05 level (.36) but not at the .01 level (.46). The obtained *r* is larger than .36 but smaller than .46. An *r* of .30 for $df = 20$ is significant at neither the .05 nor the .01 level. But an *r* of .60 for *df* of 45 is significant at *both* the .05 and .01 levels.

* The .05 and .01 levels are often written as the .95 and .99 levels of significance. The two statements say essentially the same thing. The .05 and .01 express the probability of being wrong, whereas the .95 and .99 express the probability of being right.

9.

THE CHI-SQUARE TEST

The members of a group may often be classified into subgroups or arranged into categories in terms of some observed behavior, when they cannot be measured directly in the trait itself. Classification is frequently necessary in social and clinical psychology, where many of the qualities in which the psychologist is interested (personality traits, attitudes, interests and the like) defy measurement in terms of scores. An experimenter often wishes to compare observed frequencies with frequencies to be expected on some hypothesis or in terms of some theory. The chi-square test provides a convenient method for doing this. The formula for χ^2 is

$$\chi^2 = \frac{\Sigma(o - e)^2}{e} \tag{25}$$

(chi-square formula for testing agreement between observed and expected frequencies)

in which $o =$ the observed or obtained frequencies in the various categories

$e =$ corresponding frequencies expected under some hypothesis

The difference between each observed and each expected frequency is squared, and divided by the expected or theoretical f; and the sum of these quotients is χ^2. An example will illustrate the application of the formula.

Example (1): Ten sections of Freshman English—400 students in all when combined into one group—yield the following distribution of final grades: A = 12; B = 64; C = 157; D = 100; E = 61; F = 6.

Does this distribution of marks follow the normal curve?

First we must determine the number of A's, B's, C's etc. to be expected in a group of 400 on the hypothesis of a normal distribution. Figure 24 shows how this is done. The base line of the curve is taken to cover 6σ (from 3.00σ to -3.00σ), so that each of the six grade-categories has an interval of $6\sigma/6$

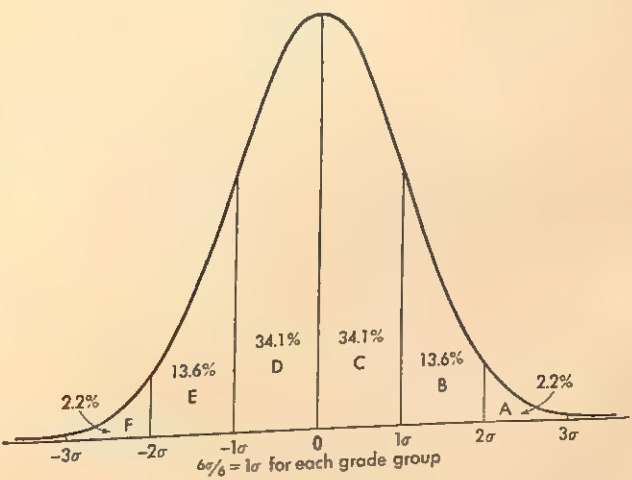


Figure 24

or 1.00σ allotted to it. The limits of each grade-category reading from right to left then become:

Grade	*Interval in* σ	*% in each category*	*Number in each category*
A	3.00 to 2.00	2.2	9
B	2.00 to 1.00	13.6	54
C	1.00 to .00	34.1	137
D	.00 to −1.00	34.1	137
E	−1.00 to −2.00	13.6	54
F	−2.00 to −3.00	2.2	9
		99.8	400

The A group extends from 3.00σ at the extreme right end of the curve to 2.00σ and from Table I we find that this portion includes 2.2% (49.87 − 47.72) of the entire area of the curve. The B group covers the interval from 2.00σ to 1.00σ and includes 13.6% (47.72 − 34.13) of the area. The C group extends from 1.00σ to .00, and embraces 34.1% of the total area. Groups D, E and F correspond exactly to C, B and A in area.

The number of students in all ten sections totals 400. We may expect, therefore, about 9 students (400 × 2.2%) to earn A's; 54 or 13.6% to earn B's; 137 or 34.1% to receive C's; 137 to

Grades

	A	B	C	D	E	F	*Total*
Observed (o)	12	64	157	100	61	6	400
Expected (e)	9	54	137	137	54	9	400
$(o-e)$	3	10	20	−37	7	−3	
$(o-e)^2$	9	100	400	1369	49	9	
$\frac{(o-e)^2}{e}$	1.00	1.85	2.92	9.99	.91	1.00	$= 17.67$ or χ^2

by (25)

Degrees of freedom $= (6-1)\,(2-1) = 5$

From Table IV, P is less than .01

receive D's; 54 to get E's and about 9 F's.* The χ^2 test enables us to compare these experimental *f*'s with those actually obtained.

As shown in the table above, each $(o - e)$ is squared and divided by its own e; and the total of the $\frac{(o-e)^2}{e} = \chi^2$. The number of degrees of freedom in the table is given by the equation:

$$df = (c - 1)\ (r - 1)$$

where $c =$ the number of columns and $r =$ the number of rows. In the present example, the $df = (6 - 1)\ (2 - 1)$ or 5.

In order to determine the significance of our χ^2 of 17.67, we must enter Table IV in the Appendix with 5 degrees of freedom and read the χ^2 values at the .05 and .01 points. At .05, χ^2 is 11.07 and at .01, χ^2 is 15.09. Since our χ^2 of 17.67 goes considerably beyond the .01 point, it is marked "significant at the .01 level." This designation means that the observed distribution of grades diverges too greatly from the expected (chance) distribution to be regarded simply as a sampling fluctuation. In other words, the *actual* distribution of grades given these 400 freshmen does *not* follow the normal curve.

A clearer notion of the meaning of levels of significance can be gotten from Figure 25 which shows the χ^2 distributions for 1, 4, 5 and 10 degrees of freedom. Consider the χ^2 curve for 5 *df*, the number of *df* in the present problem. Beginning at 0, this curve (a positively skewed distribution) runs out slightly beyond 15 where it approaches the base line more and more closely. From Table IV we read that for 5 *df*, 5% of the area of our χ^2 curve lies to the *right* of 11.07 and 1% lies to the *right* of 15.09. A χ^2 $(df = 5)$ larger than 11.07 is significant at the .05 level, and a χ^2 $(df = 5)$ larger than 15.09

* The two middle entries have been adjusted in order that the total may equal 400 exactly.

is significant at the .01 level. Only once in 100 repetitions of experiments like the one here described would we expect to find a χ^2 greater than 15.09 *if* the true value of χ^2 were zero. Any value in the region of the curve beyond 15.09 represents, therefore, an unusual value in the sense of being a very infrequent deviation from 0. Our χ^2 of 17.67 refutes the null hypothesis, and permits us to reject it with great confidence at the .01 level of significance. Knowing the *df* in a table, we can immediately test an obtained χ^2 against the two theoretical points, .05 and .01. Our confidence in the significance of a χ^2 (that is, our willingness to accept or reject the null hypothesis) depends upon whether or not χ^2 exceeds or fails to reach the .05 or .01 point. If a χ^2 fails to reach the .05 value, it is usually taken to be inconsequential and the null hypothesis is accepted. If χ^2 reaches the .05 point but fails to reach the .01 point, it may still be marked "non-significant" if we have decided *beforehand* to take the .01 value as our standard.

The χ^2 distribution for 5 *df* is one of many χ^2-curves. As the

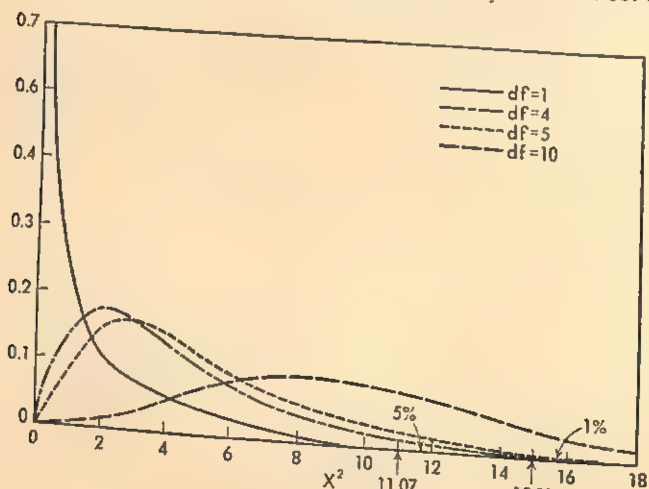


Figure 25

df increase (see Figure 25) the curves gradually lose their positive skewness and become more and more nearly normal. For $df = 1$, the curve is steeply skewed.

χ^2 *in a* 2×2 *Table*

Suppose we find that in a sample of 50 students, 30 favor and 20 oppose a given policy proposed by the class. We wish to test this result against the hypothesis that opinion is really divided equally in the group: 25 vs. 25. χ^2 is found from the table below.

	Favor	*Oppose*	*Total*
Observed (o)	30	20	50
Expected (e)	25	25	50
$(o - e)$	5	−5	
$(o - e)^2$	25	25	
$\frac{(o-e)^2}{e}$	1	1	

$\chi^2 = 2.00$ $\qquad df = 1 \qquad$ P greater than .05

The χ^2 is 2 and the $df = (2\text{-}1)(2\text{-}1)$ or 1. Entering Table IV with 1 df we find that our obtained χ^2 does not reach the .05 point of 3.84. The result is not significant, therefore, and the observed division of opinion into a (30 − 20) split could well be a chance deviation from (25 − 25). In 2×2 tables, χ^2 can be calculated somewhat more readily by the following formula, in which the expected frequencies are the same for both categories.

$$\chi^2 = \frac{2(o-e)^2}{e} \tag{26}$$

(chi-square from 2×2 table in which expected f's are the same)

In the problem above, χ^2 by formula (26) is $\frac{2 \times 5^2}{25}$ or 2.

If we had interviewed 100 students instead of 50, and opinion had divided in the same proportion as found for the smaller group (namely, 3:2), we would have found 60 in favor and 40 opposed to the given policy. The χ^2 table would then be as follows:

	Favor	*Opposed*	*Total*
o	60	40	100
e	50	50	100
$(o-e)$	10	−10	
$(o-e)^2$	100	100	
$\frac{(o-e)^2}{e}$	2	2	

$\chi^2 = 4$ $\quad df = 1$ $\quad$ *P* lies between .05 and .01

χ^2 is now 4, but the *df* is still 1. Hence, χ^2 is now significant at the .05 level (i.e., greater than 3.84) whereas the χ^2 for the sample of 50 was not significant. This result shows clearly that the larger the sample the greater the probability of a significant χ^2 (i.e., of a significant deviation from expectation) provided the ratio of *pros* and *cons* persists in the larger group. This situation is analogous to that found in tossing a coin to determine whether it is biased or off center. Getting 14 heads and 6 tails in 20 tosses of a coin is not as significant as getting 140 heads and 60 tails in 200 tosses. When the factors which produce a divergence from chance (50-50) in 20 tosses persist over 200 tosses they are not likely to be negligible: as is revealed by χ^2.

Testing Differences between Groups by Means of χ^2

χ^2 may often be used to good advantage in testing whether two or more groups differ significantly with respect to some activity, when the preferences of the groups are classified (in

terms of frequencies) into several categories. The following is an illustration.

> *Example* (1): In a study of reading preferences in 384 high-school boys, students in three courses of study (Arts, Science, Commerce) were asked to fill out a questionnaire in which they listed their preferences in books, magazines and newspapers over the past six months. Reading materials were classified under three headings and each boy placed in a category which described his *major* reading interest. Do the boys in the three courses of study differ in their reading preferences?

In order to test the null hypothesis, namely, that boys in the three curricula do not really differ in their reading preferences, we must first find the entries expected in each category under the null hypothesis. This is done in Table 17 by computing "independence values" for each classification: by finding how many boys can be expected to choose science, for example, in the absence of any real preference for this field. Independence or expected values are shown under "A" and are computed as follows. There are 114 boys who read science primarily and 149 boys enrolled in the arts curriculum. The proportion of science-reading boys in our sample, therefore, is 114/384 or 29.69%. This percentage of science readers should hold for *any* sample of high-school boys *if* there is no relation between preference for science and being in the arts curriculum. Hence $114/384 \times 149$ or 44.23 science readers should be found in the arts curriculum "by chance" alone: in the absence of any true relationship. Each expected value or independence value in "A" has been found in the manner described. A practical rule is to multiply the subtotals for a given column and row and divide by the over-all total.

TABLE 17

	Science	*Current Events*	*Fiction*	*Total*
Arts	28(44.23)	61(62.08)	60(42.68)	149
Science	52(34.44)	42(48.33)	22(33.23)	116
Commerce	34(35.33)	57(49.58)	28(34.09)	119
	114	160	110	384

A. Computation of expected (independence) values (e's)

$$\frac{114 \times 149}{384} = 44.23 \quad \frac{160 \times 149}{384} = 62.08 \quad \frac{110 \times 149}{384} = 42.68$$
$$\frac{114 \times 116}{384} = 34.44 \quad \frac{160 \times 116}{384} = 48.33 \quad \frac{110 \times 116}{384} = 33.23$$
$$\frac{114 \times 119}{384} = 35.33 \quad \frac{160 \times 119}{384} = 49.58 \quad \frac{110 \times 119}{384} = 34.09$$

B. Computation of $\frac{(o - e)^2}{e}$ values

$(28 - 44.23)^2 \div 44.23 = 5.96$ $\qquad (61 - 62.08)^2 \div 62.08 = .02$
$(52 - 34.44)^2 \div 34.44 = 8.95$ $\qquad (42 - 48.33)^2 \div 48.33 = .83$
$(34 - 35.33)^2 \div 35.33 = .05$ $\qquad (57 - 49.58)^2 \div 49.58 = 1.11$
$(60 - 42.68)^2 \div 42.68 = 7.03$
$(22 - 33.23)^2 \div 33.23 = 3.80$
$(28 - 34.09)^2 \div 34.09 = 1.09$

$\chi^2 = \frac{\Sigma(o\text{-}e)^2}{e} = 28.84$ by (25)
$df = (3\text{-}1)\ (3\text{-}1) = 4$
P is less than .01 (Table IV)

The 9 independence values found in "A" have been entered in their appropriate cells in parentheses.

In section "B," each (o-e) value—difference between observed and expected entry—has been squared and divided by the appropriate e (expected value). The sum of the $\frac{(o\text{-}e)^2}{e}$

is χ^2. In Table 17, χ^2 is 28.84 and the $df = (3\text{-}1)\ (3\text{-}1)$ or 4. Now from Table IV we find that the obtained χ^2 is far beyond the .01 point of 13.28 and hence is very significant. There is less than one chance in 100 that the divergences in reading preferences from just no preference can be explained as sampling fluctuations. We reject the null hypothesis, therefore, with great confidence and conclude that high-school boys in the three curricula do in fact differ sharply in reading preferences.

It is possible to go a step beyond the general conclusion of a positive association between curricula and reading. If we examine the $\frac{(o\text{-}e)^2}{e}$ values in "B," certain cells are seen to contribute larger amounts to χ^2 than others. As we might have surmised, boys in science read significantly more science than expectation, boys in arts definitely prefer fiction and boys in commerce prefer current events. In several cells (Commerce-Science, Arts-Current Events, for example) the divergences between obtained and expected f's are small and inconsequential.

Cautions to Be Observed in Using χ^2

There are certain restrictions to the general use of χ^2 which should be carefully observed when applying this technique. The major limitations to the χ^2 test are listed below.

(1) χ^2 is computed from a table of frequencies; it is *not* applicable to test scores. The numbers of infants who reach for an object at age 2, age 3, age 4, may be compared with the numbers who may be expected to reach for an object at each of these ages—on the basis of past standards. Or the number of "normals" who give certain answers to a questionnaire may be compared with the number of neurotics who give the

same answers. But we cannot use χ^2 to compare the *scores* made by one group with the scores made by another group.

(2) The expected or theoretical f in any cell should be at least 5 if we are to get a valid χ^2. This is a practical requirement made necessary by the fact that the χ^2 formula involves certain mathematical approximations which are not fulfilled when the expected f's are small. In 2×2 tables, when the cell entries are small, a more precise χ^2 is obtained by subtracting .5 from each of the two $(o - e)$'s. For a discussion of this correction (called Yates' correction) see more advanced texts.

(3) The observed and expected f's should add up to the same total. In Example (1) (p. 123) the 400 o's are matched against 400 e's. In Example (1) (p. 129) the sum of the obtained entries and the sum of the theoretical entries are the same, namely 384.

(4) The categories or classes into which the observed f's are placed should be independent and not overlapping. In Example (1) (p. 129) the classification would be in error if the same boy were put under both Arts and Sciences; or if two boys collaborated in their answers. The observed f in our classification is tested against the assumption of complete lack of relationship between categories. The null hypothesis is not fairly tested unless the independence values are actually what they are represented to be.

10.

COMPARING AND COMBINING TEST SCORES

Workers with mental tests frequently want (1) to compare the scores made by an individual on two or more tests, or by two individuals upon the same test, and (2) to combine the scores from separate tests into a composite which will represent general achievement. Several difficulties arise when we compare or combine scores taken from different sorts of tests. For one thing, scores upon mental tests differ widely in the *kind* of units and in the *sizes* of the units in which performance is expressed. In a vocabulary test, for example, the test unit is a *word*, in an arithmetic reasoning test it is usually a *problem*. Power tests, in which the items get progressively harder, are scored according to the number of items done correctly, time being a relatively minor factor. Speed tests, on the other hand, are scored by time-taken-to-complete, or number of items (usually easy) done in a certain time-limit. A score of 22 on a difficult analogies test cannot be compared directly with a score of 22 on an easy analogies test; and a score of 20 minutes-to-complete cannot be compared with a score of 20 items done in unlimited time. Finally, tests differ greatly in length, difficulty, range and variability.

Test scores expressed in the same *kind* of unit (for example, number of items marked) even though only roughly comparable, are sometimes combined to give an aggregate or composite score. The final score on many educational achievement tests, for example, is found by simply adding the subtest scores. The easiest procedure in combining test scores is to add them as they stand. This scheme, however, gives us no control over the relative importance or "weights" to be attached to the various subtests in the composite. It is often mistakenly assumed that by simply adding or averaging test scores we avoid the troublesome question of weighting. But what we actually do in such cases is to weight quite drastically, without knowing, however, what the weights are. Tests which are not weighted, weight themselves.

Three methods of reducing test scores to a comparable basis, so that they may be compared or combined with known weights, will be described in the present chapter.

Converting the Scores of Different Tests into Standard Deviation Units

(1) σ-SCORES OR z-SCORES

When the deviation (x) of a score (X) from the mean (M) is divided by the σ of the test, the result is known alternately as a sigma-score, or a z-score.* The formula for a sigma-score is

$$z\text{- or } \sigma\text{-score} = \frac{x}{\sigma} = \frac{X-M}{\sigma} \qquad (27)$$

(sigma-score or z-score, i.e., a test score expressed in σ-units)

Derived scores of this sort are especially useful when tests scaled in different units are to be combined into a composite. The following example will make this point clear.

* See also p. 76. x/σ deviations in the normal curve are also σ-scores.

Example (1): The M's and σ's of five tests of educational achievement are given below, together with the scores earned by two pupils, John and Mary. Combine John's and Mary's scores into a composite in which each subtest will have the same weight.

	Arithmetic computation (1)	*Arithmetic reasoning* (2)	*Reading* (3)	*Grammar* (4)	*Spelling* (5)	*Total* (sum)	*M*
Mean	62	24	137	81	21		
SD	12	7	20	16	5		
John's scores:	68	27	127	75	15		
Mary's scores:	48	18	157	89	28		
John's σ-scores:	.50	.43	−.50	−.37	−1.20	−1.14	−.23
Mary's σ-scores:	−1.17	−.86	1.00	.50	1.40	.87	.17

Each σ-score has been found by formula (27). In arithmetic computation, for example, John's σ-score is $\frac{68-62}{12}$ or .50; in arithmetic reasoning, $\frac{27-24}{7} = .43$ and so on. Mary's σ-score in arithmetic reasoning is $\frac{18-24}{7}$ or −.86; and in reading, $\frac{157-137}{20}$ or 1.00.

Since the deviation of a score which equals the M of a test is of necessity zero ($X - M = 0$), the M of a σ-score distribution is *always* .00; and the *SD* is always 1.00, since σ is the measuring unit. The composite score for John is −1.14 and for Mary .87. John is below the M on the three language tests (minus) and his composite is negative; Mary is below the M in the two arithmetic tests, but her scores are above the mean in the three language tests, so that her composite score is plus. The M of John's five σ-scores is −.23 (−1.14/5) and of Mary's .17 (.87/5).

By converting each test score into a σ-score—putting it into a distribution with $M = 0$ and $\sigma = 1.00$— we give each test

the *same weight* in the composite and equal weight in determining the M of the sigma-scores. If John's scores and Mary's scores had simply been added as they stand, the composite or the M would have been more heavily weighted for reading and grammar than for spelling and arithmetic.* Sigma-scores since they are expressed in the same unit may be added or averaged, as one prefers.

(2) STANDARD SCORES

The chief disadvantage of σ-scores is the fact that they are usually small decimals and are about as often $-$ as $+$ (below as above the M). For this reason σ-scores are usually transformed into standard scores, that is, are converted into a new distribution with the M set arbitrarily at 50 or 100, say, and SD at 10 or 20.† Suppose we decide to convert John's and Mary's σ-scores into a new distribution with M of 100 and SD of 20. The result is shown below.

TABLE 18

	Arithmetic computation (1)	*Arithmetic reasoning* (2)	*Reading* (3)	*Grammar* (4)	*Spelling* (5)	*Total*	*Mean*
John:	110	109	90	93	76	478	96
Mary:	77	83	120	110	128	518	104

These "new scores" (standard scores) may be calculated directly from the σ-scores. Thus, in arithmetic computation John's σ-score of .50 means that he is ½ SD above the M, and ½ of 20, the new SD, puts him 10 points above the new mean

* The weights of the scores entering into a composite depend upon their absolute size and upon the variability (SD) of the scores themselves.

† Other choices are $M = 500$ and $SD = 100$; $M = 50$ and $SD = 14$; $M = 10$ and $SD = 3$.

of 100, or at 110. In spelling, John's σ-score is -1.20, i.e., 1.20σ below the M. Hence, -1.20×20 (the new SD) shows that John is 24 points *below* the new M of 100 or at 76. An easy formula for converting raw test scores directly into standard scores is the following:

$$X' = \frac{\sigma'}{\sigma}(X - M) + M' \tag{28}$$

(equation for converting raw scores into standard scores with any designated M and σ)

in which

X' = the *new* or standard score
X = original or raw score
σ' = SD of the standard score distribution
σ = SD of the given test (obtained scores) distribution
M' = M of the standard score distribution
M = M of the test

To illustrate the use of formula (28) the following substitutions are made for the spelling test in Example (1) if we wish to convert the test scores into a standard score distribution with $M = 100$ and $SD = 20$.

$$X' = \frac{20}{5}(X - 21) + 100$$
$$= 4X + 16$$

For Mary's test score of 28, we have

$$X' = 4 \times 28 + 16 = 128$$

For John's test score of 15,

$$X' = 4 \times 15 + 16 = 76$$

A separate equation must be set up, of course, for each of the five tests. From these equations, it is possible to convert raw

scores directly and quickly into the new standard score distribution. When many scores are to be converted, use of formula (28) is much easier than direct calculation by way of σ-scores.

The standard scores in Table 18 are easier to comprehend and to work with than are σ-scores. At a glance we see that John is above the M of 100 in the two arithmetic tests, below in reading, grammar and spelling. Mary, on the other hand, is excellent in reading and spelling, good in grammar, and poor in the two arithmetic tests. John's general average of 96 shows him to be slightly below the M of 100; and Mary's general average of 104 shows her to be slightly above the mean in educational achievement.

Standard scores can be added, averaged and combined with equal weight. But standard scores are not strictly comparable—or equivalent *—unless the original test scores have the same *form* of distribution. Why this is true may be seen from Figures 26 and 27. The off-center distribution shown in Figure 26 is skewed or drawn out toward the *right*. If Tom has a σ-score of 1.00 (standard score 1σ above the M) in the test represented by this distribution, his score is exceeded by about 24% of the group. Now consider the off-center distribution in Figure 27, which is skewed to the *left* to the same degree that the distribution in Figure 26 is skewed to the right. If Tom achieves a σ-score of 1.00 (standard score 1σ above the M) in the test represented by this distribution, his score is exceeded by only 1% of the group. Clearly, these two identical σ-scores do not represent the same level of performance with respect to group achievement.

Fortunately, the distributions of test scores are rarely as badly skewed as are those of Figures 26 and 27. In fact, good tests given to large groups regularly return sensibly normal distributions. Because of this, we may usually compare and combine standard scores with little error. Thus, from Table 18

* See p. 140.

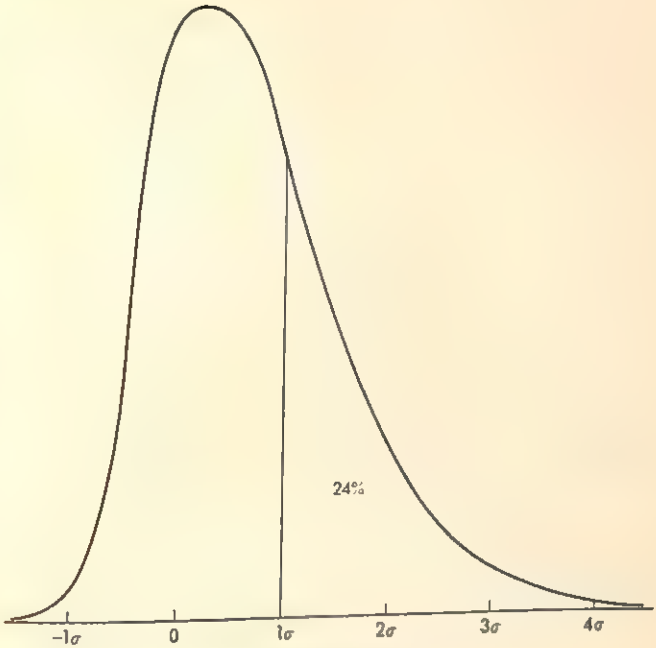


Figure 26

on page 136, we may say that John did about as well in arithmetic computation as he did in arithmetic reasoning (110 vs. 109) and that Mary did much better than John in spelling (128 vs. 76). As we shall see later (p. 143) scores *are* strictly comparable (*and* equivalent) when expressed as *T*-scores.

Converting the Scores of Different Tests into Percentile Ranks

A person who achieves a certain score on a test may be assigned a percentile rank (*PR*) of 31, 52 or 87 depending upon his position in the score distribution (p. 63). The *PR* locates a person on a scale of 100 points and tells us at once what per cent of the group scored *below* him. Furthermore, when a subject has taken a battery of tests, a comparison of

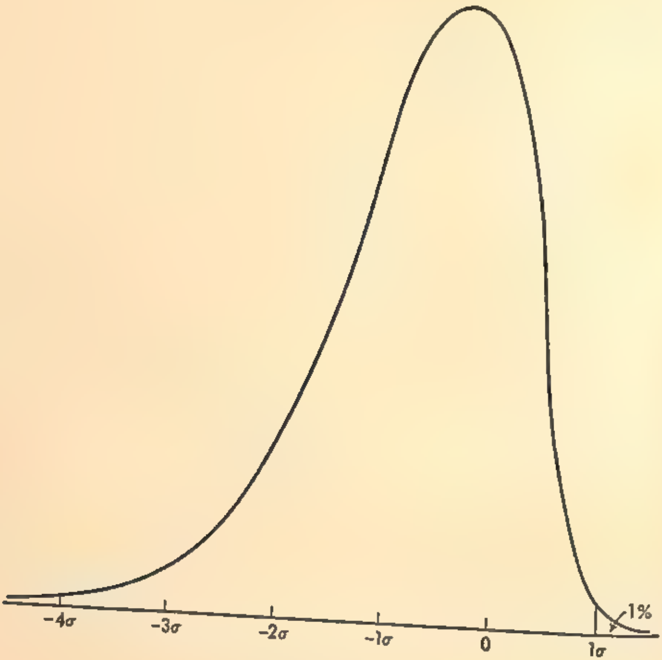


Figure 27

his *PR*'s on the subtests provides useful information concerning relative achievement.* We may also combine a person's *PR*'s into a final total score. Let us suppose that Richard has obtained the following raw scores upon five tests: general intelligence, 88; mechanical ability, 62; clerical ability, 122; arithmetic, 37; reading, 24. We can tell very little about Richard's relative achievement from his raw test scores as the test units are not comparable. Suppose, however, that Richard's *PR*'s are: general intelligence, 64; mechanical ability, 79; clerical ability, 42; arithmetic, 75; and reading, 50. We

* Scores are equivalent when they represent the same level of achievement. Thus, if a child's *PR* is 82 in reading and 82 in arithmetic, he is as good in reading as he is in arithmetic: the two scores represent equivalent levels of performance with respect to the group.

now know that in intelligence Richard is considerably above the mean (of 50) for boys of his age, is very good in mechanical ability, is good in arithmetic, average in reading and fairly poor in clerical ability. This boy's mean percentile rank is 62 which puts him 12 points above the *M* on the test battery as a whole.

The percentile scale possesses one real disadvantage, namely, that differences in *PR*'s are equal only when the distribution of test scores is *rectangular* in form. *PR* differences are *not* equal when the distribution is bell-shaped or normal.

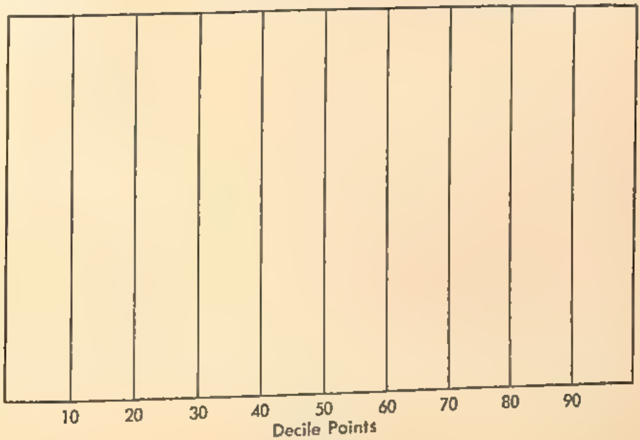


Figure 28

Figures 28 and 29 show why this is true. In Figure 28 a rectangular distribution has been divided into ten equal segments. The points along the base line represent percentile or decile (10ths) points. Note that the small rectangles are equal in size, and that the distances allotted to each tenth of the distribution along the base line are also equal. Distributions of test scores, however, tend to be normal or nearly normal and are rarely if ever truly rectangular. When the *area* of the normal curve is cut up into ten equal segments, distances

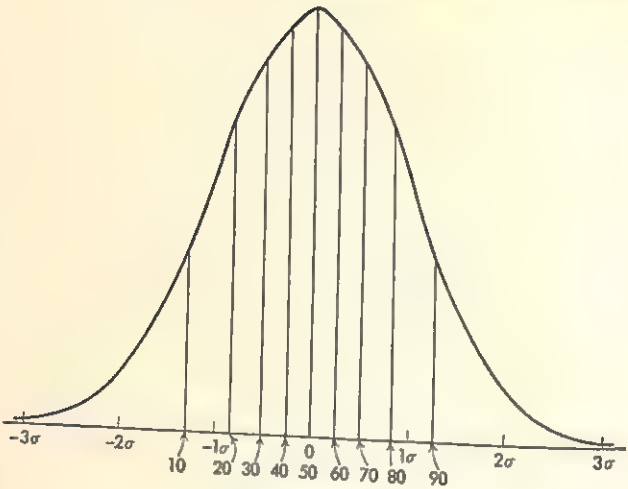


Figure 29

along the base line are no longer equal. In Figure 29, for instance, it is clear that the distance from the low end of the distribution (from -3σ) to the *PR* of 10 is very much greater (about four times) than is the distance between the *PR*'s at 40 and 50 or 50 and 60; and the distance from *PR* 90 to the upper limit of the distribution is likewise much greater than are the distances between *PR*'s at the middle of the curve. From *PR* 20 to *PR* 80, however, the distances along the base line of the normal curve are substantially equal. No great error is made, therefore, when we average or combine *PR*'s between 20 and 80. But *PR*'s above or below these boundary values should be combined, if at all, with full knowledge of their limitations.

Converting Raw or Obtained Scores into Equivalent Scores in a Normal Distribution: T-scores

The obtained (raw) scores of a frequency distribution may be converted into a system of "normalized" scores by trans-

forming them into equivalent scores in a normal distribution. Equivalent scores, in this sense, are scores which reflect the *same* level of talent or ability. Suppose that William's scores in arithmetic (28) and in history (64) are excelled by just 20% of the group in each case. From Table I (Appendix) we know that just 20% of the area in the normal curve lies to the *right* of $.84\sigma$ (30% falls between the M and $+ .84\sigma$). Both of William's scores, therefore, are "equivalent" to a "score" of $.84\sigma$ in a normal distribution, and both scores denote the same level of superiority. Normalized scores (called T-scores) differ from σ-scores and standard scores (see p. 134). The σ-score expresses the deviation of a raw score from the M in terms of σ without changing in any way the *form* of the distribution, whether normal or skewed. Standard scores are comparable (p. 138) *only* when the distributions from which they come are of the *same* form. T-scores, on the other hand, are equivalent scores found by converting the distributions of raw scores into a common normal distribution, with an M of 50 and a σ of 10. Each raw score, when transformed into a T-score, occupies the same relative position in the "standard" normal distribution as the raw score did in its own distribution. Raw scores derived from different distributions are always *equivalent* when their T-scores are equal.

The following example will show how obtained scores are transformed into T-scores.

The midpoints (column 2) best represent *all* of the scores on an interval and will hereafter stand for the separate scores on the interval. Frequencies are listed in column (3), and in column (4) these frequencies have been cumulated from the low end of the distribution upward. Column (5) is headed "*cum f* below (the midpoint) $+ \frac{1}{2}$ of the *f on* that midpoint." For example, 126 *cum f*'s [column (4)] take us up to 137.5, lower limit of interval (138-140). Adding to 126 one-half of the *f* on (138-140), namely, ½ of 21 or 10.5, we have 136.5 *f*'s

Example (1): Given a distribution of 200 aptitude test scores. Transform these raw scores into *T*-scores.

Intervals (1)	*Midpoints* (2)	*f* (3)	*cum f* (4)	*cum f below + ½ on midpoint* (5)	*cum % f below + ½ on midpoint* (6)	*T-score* (7)
147-149	148	13	200	193.5	96.7	68
144-146	145	18	187	178	89	62
141-143	142	22	169	158	79	58
138-140	139	21	147	136.5	68.3	55
135-137	136	26	126	113	56.5	52
132-134	133	40	100	80	40	47
129-131	130	28	60	46	23	43
126-128	127	20	32	22	11	38
123-125	124	12	12	6	3	31
		$N = 200$				

up to the middle of (138-140), or up to 139. The score of 139 now constitutes a *point* in the distribution below which 136.5 or 68.3% of the total *N* lies. In like manner, the midpoint of score 142 is a point below which 158 (147 + ½ of 22) or 79% of the *f* falls, and so on. In column (6), each of the entries in column (5) is turned into a percentage by dividing by *N* (200).

By means of column (6) and Table V (Appendix) we may

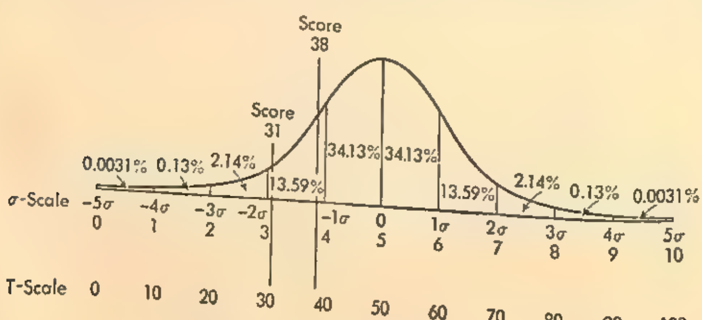


Figure 30

now transform our raw scores into the T-scores (normalized scores) shown in column (7). The normal distribution pictured in Figure 30 will serve as a model to show how this is done. In Figure 30 the base line has been marked off into ten equal σ-divisions: 5σ to the right and 5σ to the left of the M. The point at -5σ is then called 0, -4σ becomes 10, and -3σ is marked 20 and so on to the middle of the curve which is 50. Above 50, the midpoint, we have 60, 70, 80, 90, and 100 at 1, 2, 3, 4, 5, σ-intervals. This is our model distribution of T-scores; the mean is set at 50 and the σ is 10; the base line scale stretches from 0 to 100.

This scale of 100 points constitutes the T-scale into which our raw scores are converted. Table V permits us to make conversions from raw scores to T-scores quite easily. From Table V, for example, we find that 3% of area from the left end of the normal curve takes us to a T-score of 31; that 11% of area from the left end of the curve yields a T-score of 38; 23%, a T-score of 43 and so on. The first two T-scores are marked off in Figure 30.*

T-scores or normalized scores are useful in enabling us to compare and combine test scores expressed in different units. Thus T-scores of 65 in mechanical comprehension and 65 in arithmetic reasoning express exactly the same degree of achievement relative to the group performance.

T-scores are superior to standard scores (which they resemble superficially) because no question about distribution form arises in making comparisons. Two standard scores are comparable *only* when the raw scores which they represent have similar distributions (normal, or skewed in exactly the

* Note that these T-scores may also be read from Table I. Thus the lowest 3% of the normal distribution falls below 1.9σ (i.e., -1.9σ from the mean). In the T distribution the σ is 10 so that -1.9σ becomes -19 from the M of 50 or at 31. Also, the lowest 11% of the normal distribution falls below 1.2σ. This point is -1.2σ from the M of the normal distribution or -12 from the M of 50 in the T distribution (i.e., at 38).

same fashion, p. 138). Equal T-scores, on the other hand, are *always* equivalent since they represent scores converted into a common model (normal) distribution. Suppose that John has earned scores of 43 in arithmetic, 62 in reading, 81 in history and 34 in spelling. Suppose further that the distributions of these scores are normalized (raw scores translated into T-scores) and John now has the following T-scores: arithmetic 54, reading 47, history 73, and spelling 44. These scores tell us at once how John stands with reference to the mean (50) of his age or grade. Furthermore, if John has T-scores of 62 in science and 62 in geography he is (with respect to the group) as good in the one subject as he is in the other.

TABLES IN THE APPENDIX

TABLE I

Normal Probability Curve

Per Cent of Total Area under the Normal Curve between Mean Ordinate And Ordinate at Any Given Sigma-Distance from the Mean

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	00.00	00.40	00.80	01.20	01.60	01.99	02.39	02.79	03.19	03.59
0.1	03.98	04.38	04.78	05.17	05.57	05.96	06.36	06.75	07.14	07.53
0.2	07.93	08.32	08.71	09.10	09.48	09.87	10.26	10.64	11.03	11.41
0.3	11.79	12.17	12.55	12.93	13.31	13.68	14.06	14.43	14.80	15.17
0.4	15.54	15.91	16.28	16.64	17.00	17.36	17.72	18.08	18.44	18.79
0.5	19.15	19.50	19.85	20.19	20.54	20.88	21.23	21.57	21.90	22.24
0.6	22.57	22.91	23.24	23.57	23.89	24.22	24.54	24.86	25.17	25.49
0.7	25.80	26.11	26.42	26.73	27.04	27.34	27.64	27.94	28.23	28.52
0.8	28.81	29.10	29.39	29.67	29.95	30.23	30.51	30.78	31.06	31.33
0.9	31.59	31.86	32.12	32.38	32.64	32.90	33.15	33.40	33.65	33.89
1.0	34.13	34.38	34.61	34.85	35.08	35.31	35.54	35.77	35.99	36.21
1.1	36.43	36.65	36.86	37.08	37.29	37.49	37.70	37.90	38.10	38.30
1.2	38.49	38.69	38.88	39.07	39.25	39.44	39.62	39.80	39.97	40.15
1.3	40.32	40.49	40.66	40.82	40.99	41.15	41.31	41.47	41.62	41.77
1.4	41.92	42.07	42.22	42.36	42.51	42.65	42.79	42.92	43.06	43.19
1.5	43.32	43.45	43.57	43.70	43.83	43.94	44.06	44.18	44.29	44.41
1.6	44.52	44.63	44.74	44.84	44.95	45.05	45.15	45.25	45.35	45.45
1.7	45.54	45.64	45.73	45.82	45.91	45.99	46.08	46.16	46.25	46.33
1.8	46.41	46.49	46.56	46.64	46.71	46.78	46.86	46.93	46.99	47.06
1.9	47.13	47.19	47.26	47.32	47.38	47.44	47.50	47.56	47.61	47.67
2.0	47.72	47.78	47.83	47.88	47.93	47.98	48.03	48.08	48.12	48.17
2.1	48.21	48.26	48.30	48.34	48.38	48.42	48.46	48.50	48.54	48.57
2.2	48.61	48.64	48.68	48.71	48.75	48.78	48.81	48.84	48.87	48.90
2.3	48.93	48.96	48.98	49.01	49.04	49.06	49.09	49.11	49.13	49.16
2.4	49.18	49.20	49.22	49.25	49.27	49.29	49.31	49.32	49.34	49.36
2.5	49.38	49.40	49.41	49.43	49.45	49.46	49.48	49.49	49.51	49.52
2.6	49.53	49.55	49.56	49.57	49.59	49.60	49.61	49.62	49.63	49.64
2.7	49.65	49.66	49.67	49.68	49.69	49.70	49.71	49.72	49.73	49.74
2.8	49.74	49.75	49.76	49.77	49.77	49.78	49.79	49.79	49.80	49.81
2.9	49.81	49.82	49.82	49.83	49.84	49.84	49.85	49.85	49.86	49.86
3.0	49.87									
3.5	49.98									
4.0	49.997									
5.0	49.99997									

* The original data for Table B came from, *Tables for statisticians and biometricians*, edited by Karl Pearson, published by Cambridge University Press, and are used here by permission of the publisher. The adaptation of these data is taken from Lindquist, E. L., *A first course in statistics* (revised edition), with permission of the publisher, Houghton Mifflin Company.

TABLE II

Values of t (the critical ratio) at the .05 and the .01 Levels of Significance

Example: When the *df* are 20 and the *t* is 2.09, the .05 level means that 5 times in 100 trials a divergence as large as or larger than that obtained (plus or minus) may be expected under the null hypothesis.

Degrees of freedom *df*	.05	.01
1	12.71	63.66
2	4.30	9.92
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.36	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3.17
11	2.20	3.11
12	2.18	3.06
13	2.16	3.01
14	2.14	2.98
15	2.13	2.95
16	2.12	2.92
17	2.11	2.90
18	2.10	2.88
19	2.09	2.86
20	2.09	2.84
21	2.08	2.83
22	2.07	2.82
23	2.07	2.81
24	2.06	2.80
25	2.06	2.79
26	2.06	2.78
27	2.05	2.77

Degrees of freedom *df*	.05	.01
28	2.05	2.76
29	2.04	2.76
30	2.04	2.75
.	.	.
.	.	.
.	.	.
50	2.01	2.68
.	.	.
100	1.98	2.63
.	.	.
.	.	.
Over 100	1.96	2.58

TABLE III

Values of r, the Coefficient of Correlation, at the .05 and .01 Levels of Significance

Example: When N is 30 and the df 28, an r must be as large as .36 to be significant at the 5% level, and .46 to be significant at the 1% level.

Degrees of freedom (df) (N − 2)	.05	.01	*Degrees of freedom (df)* (N − 2)	.05	.01
1	.997	1.000	24	.39	.50
2	.95	.99	25	.38	.49
3	.88	.96	26	.37	.48
4	.81	.92	27	.37	.47
5	.75	.87	28	.36	.46
6	.71	.83	29	.36	.46
7	.67	.80	30	.35	.45
8	.63	.77	35	.33	.42
9	.60	.74	40	.30	.39
10	.58	.71	45	.29	.37
11	.55	.68	50	.27	.35
12	.53	.66	60	.25	.33
13	.51	.64	70	.23	.30
14	.50	.62	80	.22	.28
15	.48	.61	90	.21	.27
16	.47	.59	100	.20	.25
17	.46	.58	125	.17	.23
18	.44	.56	150	.16	.21
19	.43	.55	200	.14	.18
20	.42	.54	300	.11	.15
21	.41	.53	400	.10	.13
22	.40	.52	500	.09	.12
23	.40	.51	1000	.06	.08

TABLE IV

Values of Chi-square (χ^2) at the .05 and the .01 Levels of Significance

Example: For 12 degrees of freedom, a computed χ^2 must be at least as large as 21.03 to be significant at the 5% level, and as large as 26.22 to be significant at the 1% level.

Degrees of freedom (*df*)	.05	.01
1	3.84	6.64
2	5.99	9.21
3	7.82	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.72
12	21.03	26.22
13	22.36	27.69
14	23.68	29.14
15	25.00	30.58
16	26.30	32.00
17	27.59	33.41
18	28.87	34.80
19	30.14	36.19
20	31.41	37.57
21	32.67	38.93
22	33.92	40.29
23	35.17	41.64
24	36.42	42.98
25	37.65	44.31
26	38.88	45.64
27	40.11	46.96
28	41.34	48.28
29	42.56	49.59
30	43.77	50.89

TABLE V

To Facilitate the Calculation of T-scores

The per cents refer to the percentage of the total frequency below a given score + ½ of the frequency on that score. *T-scores* are read directly from the given percentages.

Per cent	*T-score*	*Per cent*	*T-score*
.13	20	53.98	51
.19	21	57.93	52
.26	22	61.79	53
.35	23	65.54	54
.47	24	69.15	55
.62	25	72.57	56
.82	26	75.80	57
1.07	27	78.81	58
1.39	28	81.59	59
1.79	29	84.13	60
2.28	30	86.43	61
2.87	31	88.49	62
3.59	32	90.32	63
4.46	33	91.92	64
5.48	34	93.32	65
6.68	35	94.52	66
8.08	36	95.54	67
9.68	37	96.41	68
11.51	38	97.13	69
13.57	39	97.72	70
15.87	40	98.21	71
18.41	41	98.61	72
21.19	42	98.93	73
24.20	43	99.18	74
27.43	44	99.38	75
30.85	45	99.53	76
34.46	46	99.65	77
38.21	47	99.74	78
42.07	48	99.81	79
46.02	49	99.865	80
50.00	50		

TABLE OF SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square Root	Number	Square	Square Root
1	1	1.000	51	26 01	7.141
2	4	1.414	52	27 04	7.211
3	9	1.732	53	28 09	7.280
4	16	2.000	54	29 16	7.348
5	25	2.236	55	30 25	7.416
6	36	2.449	56	31 36	7.483
7	49	2.646	57	32 49	7.550
8	64	2.828	58	33 64	7.616
9	81	3.000	59	34 81	7.681
10	1 00	3.162	60	36 00	7.746
11	1 21	3.317	61	37 21	7.810
12	1 44	3.464	62	38 44	7.874
13	1 69	3.606	63	39 69	7.937
14	1 96	3.742	64	40 96	8.000
15	2 25	3.873	65	42 25	8 062
16	2 56	4.000	66	43 56	8.124
17	2 89	4.123	67	44 89	8.185
18	3 24	4.243	68	46 24	8.246
19	3 61	4.359	69	47 61	8.307
20	4 00	4.472	70	49 00	8.367
21	4 41	4.583	71	50 41	8.426
22	4 84	4.690	72	51 84	8.485
23	5 29	4.796	73	53 29	8.544
24	5 76	4.899	74	54 76	8.602
25	6 25	5.000	75	56 25	8.660
26	6 76	5.099	76	57 76	8.718
27	7 29	5.196	77	59 29	8.775
28	7 84	5.292	78	60 84	8.832
29	8 41	5.385	79	62 41	8.888
30	9 00	5.477	80	64 00	8.944
31	9 61	5.568	81	65 61	9.000
32	10 24	5.657	82	67 24	9.055
33	10 89	5.745	83	68 89	9.110
34	11 56	5.831	84	70 56	9.165
35	12 25	5.916	85	72 25	9.220
36	12 96	6.000	86	73 96	9.274
37	13 69	6.083	87	75 69	9.327
38	14 44	6.164	88	77 44	9.381
39	15 21	6.245	89	79 21	9.434
40	16 00	6.325	90	81 00	9.487
41	16 81	6.403	91	82 81	9.539
42	17 64	6.481	92	84 64	9.592
43	18 49	6.557	93	86 49	9.644
44	19 36	6.633	94	88 36	9.695
45	20 25	6.708	95	90 25	9.747
46	21 16	6.782	96	92 16	9.798
47	22 09	6.856	97	94 09	9.849
48	23 04	6.928	98	96 04	9 899
49	24 01	7.000	99	98 01	9 950
50	25 00	7.071	100	1 00 00	10 000

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
101	1 02 01	10.050	151	2 28 01	12.288
102	1 04 04	10.100	152	2 31 04	12.329
103	1 06 09	10.149	153	2 34 09	12.369
104	1 08 16	10.198	154	2 37 16	12.410
105	1 10 25	10.247	155	2 40 25	12.450
106	1 12 36	10.296	156	2 43 36	12.490
107	1 14 49	10.344	157	2 46 49	12.530
108	1 16 64	10.392	158	2 49 64	12.570
109	1 18 81	10.440	159	2 52 81	12.610
110	1 21 00	10.488	160	2 56 00	12.649
111	1 23 21	10.536	161	2 59 21	12.689
112	1 25 44	10.583	162	2 62 44	12.728
113	1 27.69	10.630	163	2 65 69	12.767
114	1 29 96	10.677	164	2 68 96	12.806
115	1 32 25	10.724	165	2 72 25	12.845
116	1 34 56	10.770	166	2 75 56	12.884
117	1 36 89	10.817	167	2 78 89	12.923
118	1 39 24	10.863	168	2 82 24	12.961
119	1 41 61	10.909	169	2 85 61	13.000
120	1 44 00	10.954	170	2 89 00	13.038
121	1 46 41	11.000	171	2 92 41	13.077
122	1 48 84	11.045	172	2 95 84	13.115
123	1 51 29	11.091	173	2 99 29	13.153
124	1 53 76	11.136	174	3 02 76	13.191
125	1 56 25	11.180	175	3 06 25	13.229
126	1 58 76	11.225	176	3 09 76	13.266
127	1 61 29	11.269	177	3 13 29	13 304
128	1 63 84	11.314	178	3 16 84	13.342
129	1 66 41	11 358	179	3 20 41	13.379
130	1 69 00	11.402	180	3 24 00	13.416
131	1 71 61	11.446	181	3 27 61	13.454
132	1 74 24	11.489	182	3 31 24	13.491
133	1 76 89	11.533	183	3 34 89	13.528
134	1 79 56	11.576	184	3 38 56	13.565
135	1 82 25	11.619	185	3 42 25	13.601
136	1 84 96	11.662	186	3 45 96	13.638
137	1 87 69	11.705	187	3 49 69	13.675
138	1 90 44	11.747	188	3 53 44	13.711
139	1 93 21	11.790	189	3 57 21	13.748
140	1 96 00	11.832	190	3 61 00	13.784
141	1 98 81	11.874	191	3 64 81	13.820
142	2 01 64	11.916	192	3 68 64	13.856
143	2 04 49	11.958	193	3 72 49	13.892
144	2 07 36	12.000	194	3 76 36	13.928
145	2 10 25	12.042	195	3 80 25	13.964
146	2 13 16	12.083	196	3 84 16	14.000
147	2 16 09	12.124	197	3 88 09	14.036
148	2 19 04	12.166	198	3 92 04	14.071
149	2 22 01	12.207	199	3 96 01	14.107
150	2 25 00	12.247	200	4 00 00	14.142

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
201	4 04 01	14.177	251	6 30 01	15.843
202	4 08 04	14.213	252	6 35 04	15.875
203	4 12 09	14.248	253	6 40 09	15.906
204	4 16 16	14.283	254	6 45 16	15.937
205	4 20 25	14.318	255	6 50 25	15.969
206	4 24 36	14.353	256	6 55 36	16.000
207	4 28 49	14.387	257	6 60 49	16.031
208	4 32 64	14.422	258	6 65 64	16.062
209	4 36 81	14.457	259	6 70 81	16.093
210	4 41 00	14.491	260	6 76 00	16.125
211	4 45 21	14.526	261	6 81 21	16.155
212	4 49 44	14.560	262	6 86 44	16.186
213	4 53 69	14.595	263	6 91 69	16.217
214	4 57 96	14.629	264	6 96 96	16.248
215	4 62 25	14.663	265	7 02 25	16.279
216	4 66 56	14.697	266	7 07 56	16.310
217	4 70 89	14.731	267	7·12 89	16.340
218	4 75 24	14.765	268	7 18 24	16.371
219	4 79 61	14.799	269	7 23 61	16.401
220	4 84 00	14.832	270	7 29 00	16.432
221	4 88 41	14.866	271	7 34 41	16.462
222	4 92 84	14.900	272	7 39 84	16.492
223	4 97 29	14.933	273	7 45 29	16.523
224	5 01 76	14.967	274	7 50 76	16.553
225	5 06 25	15.000	275	7 56 25	16.583
226	5 10 76	15.033	276	7 61 76	16.613
227	5 15 29	15.067	277	7 67 29	16.643
228	5 19 84	15.100	278	7 72 84	16.673
229	5 24 41	15.133	279	7 78 41	16.703
230	5 29 00	15.166	280	7 84 00	16.733
231	5 33 61	15.199	281	7 89 61	16.763
232	5 38 24	15.232	282	7 95 24	16.793
233	5 42 89	15.264	283	8 00 89	16.823
234	5 47 56	15.297	284	8 06 56	16.852
235	5 52 25	15.330	285	8 12 25	16.882
236	5 56 96	15.362	286	8 17 96	16.912
237	5 61 69	15.395	287	8 23 69	16.941
238	5 66 44	15.427	288	8 29 44	16.971
239	5 71 21	15.460	289	8 35 21	17.000
240	5 76 00	15.492	290	8 41 00	17.029
241	5 80 81	15.524	291	8 46 81	17.059
242	5 85 64	15.556	292	8 52 64	17.088
243	5 90 49	15.588	293	8 58 49	17.117
244	5 95 36	15.620	294	8 64 36	17 146
245	6 00 25	15.652	295	8 70 25	17.176
246	6 05 16	15.684	296	8 76 16	17.205
247	6 10 09	15.716	297	8 82 09	17.234
248	6 15 04	15.748	298	8 88 04	17.263
249	6 20 01	15.780	299	8 94 01	17.292
250	6 25 00	15.811	300	9 00 00	17.321

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root
301	9 06 01	17.349
302	9 12 04	17.378
303	9 18 09	17.407
304	9 24 16	17.436
305	9 30 25	17.464
306	9 36 36	17.493
307	9 42 49	17.521
308	9 48 64	17.550
309	9 54 81	17.578
310	9 61 00	17.607
311	9 67 21	17 635
312	9 73 44	17 664
313	9 79 69	17.692
314	9 85 96	17.720
315	9 92 25	17.748
316	9 98 56	17.776
317	10 04 89	17.804
318	10 11 24	17 833
319	10 17 61	17.861
320	10 24 00	17.889
321	10 30 41	17.916
322	10 36 84	17 944
323	10 43 29	17 972
324	10 49 76	18.000
325	10 56 25	18.028
326	10 62 76	18.055
327	10 69 29	18.083
328	10 75 84	18.111
329	10 82 41	18.138
330	10 89 00	18.166
331	10 95 61	18.193
332	11 02 24	18.221
333	11 08 89	18.248
334	11 15 56	18.276
335	11 22 25	18.303
336	11 28 96	18.330
337	11 35 69	18.358
338	11 42 44	18.385
339	11 49 21	18.412
340	11 56 00	18.439
341	11 62 81	18.466
342	11 69 64	18.493
343	11 76 49	18.520
344	11 83 36	18.547
345	11 90 25	18.574
346	11 97 16	18.601
347	12 04 09	18.628
348	12 11 04	18.655
349	12 18 01	18.682
350	12 25 00	18.708
351	12 32 01	18.735
352	12 39 04	18.762
353	12 46 09	18.788
354	12 53 16	18.815
355	12 60 25	18.841
356	12 67 36	18.868
357	12 74 49	18 894
358	12 81 64	18.921
359	12 88 81	18.947
360	12 96 00	18.974
361	13 03 21	19.000
362	13 10 44	19.026
363	13 17 69	19.053
364	13 24 96	19.079
365	13 32 25	19.105
366	13 39 56	19.131
367	13 46 89	19.157
368	13 54 24	19.183
369	13 61 61	19.209
370	13 69 00	19.235
371	13 76 41	19.261
372	13 83 84	19.287
373	13 91 29	19.313
374	13 98 76	19.339
375	14 06 25	19.363
376	14 13 76	19.391
377	14 21 29	19.416
378	14 28 84	19.442
379	14 36 41	19.468
380	14 44 00	19.494
381	14 51 61	19.519
382	14 59 24	19.545
383	14 66 89	19.570
384	14 74 56	19.596
385	14 82 25	19.621
386	14 89 96	19.647
387	14 97 69	19.672
388	15 05 44	19.698
389	15 13 21	19.723
390	15 21 00	19.748
391	15 28 81	19.774
392	15 36 64	19.799
393	15 44 49	19.824
394	15 52 36	19.849
395	15 60 25	19.875
396	15 68 16	19.900
397	15 76 09	19.925
398	15 84 04	19.950
399	15 92 01	19.975
400	16 00 00	20.000

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
401	16 08 01	20.025	451	20 34 01	21.237
402	16 16 04	20.050	452	20 43 04	21.260
403	16 24 09	20.075	453	20 52 09	21.284
404	16 32 16	20.100	454	20 61 16	21.307
405	16 40 25	20.125	455	20 70 25	21.331
406	16 48 36	20.149	456	20 79 36	21.354
407	16 56 49	20.174	457	20 88 49	21.378
408	16 64 64	20.199	458	20 97 64	21.401
409	16 72 81	20.224	459	21 06 81	21.424
410	16 81 00	20.248	460	21 16 00	21.448
411	16 89 21	20.273	461	21 25 21	21.471
412	16 97 44	20.298	462	21 34 44	21.494
413	17 05 69	20.322	463	21 43 69	21.517
414	17 13 96	20.347	464	21 52 96	21.541
415	17 22 25	20.372	465	21 62 25	21.564
416	17 30 56	20.396	466	21 71 56	21.587
417	17 38 89	20.421	467	21 80 89	21.610
418	17 47 24	20.445	468	21 90 24	21.633
419	17 55 61	20.469	469	21 99 61	21.656
420	17 64 00	20.494	470	22 09 00	21.679
421	17 72 41	20.518	471	22 18 41	21.703
422	17 80 84	20.543	472	22 27 84	21.726
423	17 89 29	20.567	473	22 37 29	21.749
424	17 97 76	20.591	474	22 46 76	21.772
425	18 06 25	20.616	475	22 56 25	21.794
426	18 14 76	20.640	476	22 65 76	21.817
427	18 23 29	20.664	477	22 75 29	21.840
428	18 31 84	20.688	478	22 84 84	21.863
429	18 40 41	20.712	479	22 94 41	21.886
430	18 49 00	20.736	480	23 04 00	21.909
431	18 57 61	20.761	481	23 13 61	21.932
432	18 66 24	20.785	482	23 23 24	21.954
433	18 74 89	20.809	483	23 32 89	21.977
434	18 83 56	20.833	484	23 42 56	22.000
435	18 92 25	20.857	485	23 52 25	22.023
436	19 00 96	20.881	486	23 61 96	22.045
437	19 09 69	20.905	487	23 71 69	22.068
438	19 18 44	20.928	488	23 81 44	22.091
439	19 27 21	20.952	489	23 91 21	22.113
440	19 36 00	20.976	490	24 01 00	22.136
441	19 44 81	21.000	491	24 10 81	22.159
442	19 53 64	21.024	492	24 20 64	22.181
443	19 62 49	21.048	493	24 30 49	22.204
444	19 71 36	21.071	494	24 40 36	22.226
445	19 80 25	21.095	495	24 50 25	22.249
446	19 89 16	21.119	496	24 60 16	22.271
447	19 98 09	21.142	497	24 70 09	22.293
448	20 07 04	21.166	498	24 80 04	22.316
449	20 16 01	21.190	499	24 90 01	22.338
450	20 25 00	21.213	500	25 00 00	22.361

TABLE OF SQUARES AND SQUARE ROOTS—*Continued*

Number	Square	Square Root	Number	Square	Square Root
501	25 10 01	22.383	551	30 36 01	23.473
502	25 20 04	22.405	552	30 47 04	23.495
503	25 30 09	22.428	553	30 58 09	23.516
504	25 40 16	22.450	554	30 69 16	23.537
505	25 50 25	22.472	555	30 80 25	23.558
506	25 60 36	22.494	556	30 91 36	23.580
507	25 70 49	22.517	557	31 02 49	23.601
508	25 80 64	22.539	558	31 13 64	23.622
509	25 90 81	22.561	559	31 24 81	23.643
510	26 01 00	22.583	560	31 36 00	23.664
511	26 11 21	22.605	561	31 47 21	23.685
512	26 21 44	22.627	562	31 58 44	23.707
513	26 31 69	22.650	563	31 69 69	23.728
514	26 41 96	22.672	564	31 80 96	23.749
515	26 52 25	22.694	565	31 92 25	23.770
516	26 62 56	22.716	566	32 03 56	23.791
517	26 72 89	22.738	567	32 14 89	23.812
518	26 83 24	22.760	568	32 26 24	23.833
519	26 93 61	22.782	569	32 37 61	23.854
520	27 04 00	22.804	570	32 49 00	23.875
521	27 14 41	22.825	571	32 60 41	23.896
522	27 24 84	22.847	572	32 71 84	23.917
523	27 35 29	22.869	573	32 83 29	23.937
524	27 45 76	22.891	574	32 94 76	23.958
525	27 56 25	22.913	575	33 06 25	23.979
526	27 66 76	22.935	576	33 17 76	24.000
527	27 77 29	22.956	577	33 29 29	24.021
528	27 87 84	22.978	578	33 40 84	24.042
529	27 98 41	23.000	579	33 52 41	24.062
530	28 09 00	23.022	580	33 64 00	24.083
531	28 19 61	23.043	581	33 75 61	24.104
532	28 30 24	23.065	582	33 87 24	24.125
533	28 40 89	23.087	583	33 98 89	24.145
534	28 51 56	23.108	584	34 10 56	24.166
535	28 62 25	23.130	585	34 22 25	24.187
536	28 72 96	23.152	586	34 33 96	24.207
537	28 83 69	23.173	587	34 45 69	24.228
538	28 94 44	23.195	588	34 57 44	24.249
539	29 05 21	23.216	589	34 69 21	24.269
540	29 16 00	23.238	590	34 81 00	24.290
541	29 26 81	23.259	591	34 92 81	24.310
542	29 37 64	23.281	592	35 04 64	24.331
543	29 48 49	23.302	593	35 16 49	24.352
544	29 59 36	23.324	594	35 28 36	24.372
545	29 70 25	23.345	595	35 40 25	24.393
546	29 81 16	23.367	596	35 52 16	24.413
547	29 92 09	23.388	597	35 64 09	24.434
548	30 03 04	23.409	598	35 76 04	24.454
549	30 14 01	23.431	599	35 88 01	24.474
550	30 25 00	23.452	600	36 00 00	24.495

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
601	36 12 01	24.515	651	42 38 01	25.515
602	36 24 04	24.536	652	42 51 04	25.534
603	36 36 09	24.556	653	42 64 09	25.554
604	36 48 16	24.576	654	42 77 16	25.573
605	36 60 25	24.597	655	42 90 25	25.593
606	36 72 36	24.617	656	43 03 36	25.612
607	36 84 49	24.637	657	43 16 49	25.632
608	36 96 64	24.658	658	43 29 64	25.652
609	37 08 81	24.678	659	43 42 81	25.671
610	37 21 00	24.698	660	43 56 00	25.690
611	37 33 21	24.718	661	43 69 21	25.710
612	37 45 44	24.739	662	43 82 44	25.729
613	37 57 69	24.759	663	43 95 69	25.749
614	37 69 96	24.779	664	44 08 96	25.768
615	37 82 25	24.799	665	44 22 25	25.788
616	37 94 56	24.819	666	44 35 56	25.807
617	38 06 89	24.839	667	44 48 89	25.826
618	38 19 24	24.860	668	44 62 24	25.846
619	38 31 61	24.880	669	44 75 61	25.865
620	38 44 00	24.900	670	44 89 00	25.884
621	38 56 41	24.920	671	45 02 41	25.904
622	38 68 84	24.940	672	45 15 84	25.923
623	38 81 29	24.960	673	45 29 29	25.942
624	38 93 76	24.980	674	45 42 76	25.962
625	39 06 25	25.000	675	45 56 25	25.981
626	39 18 76	25.020	676	45 69 76	26.000
627	39 31 29	25.040	677	45 83 29	26.019
628	39 43 84	25.060	678	45 96 84	26.038
629	39 56 41	25.080	679	46 10 41	26.058
630	39 69 00	25.100	680	46 24 00	26.077
631	39 81 61	25.120	681	46 37 61	26.096
632	39 94 24	25.140	682	46 51 24	26.115
633	40 06 89	25.159	683	46 64 89	26.134
634	40 19 56	25.179	684	46 78 56	26.153
635	40 32 25	25.199	685	46 92 25	26.173
636	40 44 96	25.219	686	47 05 96	26.192
637	40 57 69	25.239	687	47 19 69	26.211
638	40 70 44	25.259	688	47 33 44	26.230
639	40 83 21	25.278	689	47 47 21	26.249
640	40 96 00	25.298	690	47 61 00	26.268
641	41 08 81	25.318	691	47 74 81	26.287
642	41 21 64	25.338	692	47 88 64	26.306
643	41 34 49	25.357	693	48 02 49	26.325
644	41 47 36	25.377	694	48 16 36	26.344
645	41 60 25	25.397	695	48 30 25	26.363
646	41 73 16	25.417	696	48 44 16	26.382
647	41 86 09	25.436	697	48 58 09	26.401
648	41 99 04	25.456	698	48 72 04	26.420
649	42 12 01	25.475	699	48 86 01	26.439
650	42 25 00	25.495	700	49 00 00	26.458

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
701	49 14 01	26.476	751	56 40 01	27.404
702	49 28 04	26.495	752	56 55 04	27.423
703	49 42 09	26.514	753	56 70 09	27.441
704	49 56 16	26.533	754	56 85 16	27.459
705	49 70 25	26.552	755	57 00 25	27.477
706	49 84 36	26.571	756	57 15 36	27.495
707	49 98 49	26.589	757	57 30 49	27.514
708	50 12 64	26.608	758	57 45 64	27.532
709	50 26 81	26.627	759	57 60 81	27.550
710	50 41 00	26.646	760	57 76 00	27.568
711	50 55 21	26.665	761	57 91 21	27.586
712	50 69 44	26.683	762	58 06 44	27.604
713	50 83 69	26.702	763	58 21 69	27.622
714	50 97 96	26.721	764	58 36 96	27.641
715	51 12 25	26.739	765	58 52 25	27.659
716	51 26 56	26.758	766	58 67 56	27.677
717	51 40 89	26.777	767	58 82 89	27.695
718	51 55 24	26.796	768	58 98 24	27.713
719	51 69 61	26.814	769	59 13 61	27.731
720	51 84 00	26.833	770	59 29 00	27.749
721	51 98 41	26.851	771	59 44 41	27.767
722	52 12 84	26.870	772	59 59 84	27.785
723	52 27 29	26.889	773	59 75 29	27.803
724	52 41 76	26.907	774	59 90 76	27.821
725	52 56 25	26.926	775	60 06 25	27.839
726	52 70 76	26.944	776	60 21 76	27.857
727	52 85 29	26.963	777	60 37 29	27.875
728	52 99 84	26.981	778	60 52 84	27.893
729	53 14 41	27.000	779	60 68 41	27.911
730	53 29 00	27.019	780	60 84 00	27.928
731	53 43 61	27.037	781	60 99 61	27.946
732	53 58 24	27.055	782	61 15 24	27.964
733	53 72 89	27.074	783	61 30 89	27.982
734	53 87 56	27.092	784	61 46 56	28.000
735	54 02 25	27.111	785	61 62 25	28.018
736	54 16 96	27.129	786	61 77 96	28.036
737	54 31 69	27.148	787	61 93 69	28.054
738	54 46 44	27.166	788	62 09 44	28.071
739	54 61 21	27.185	789	62 25 21	28.089
740	54 76 00	27.203	790	62 41 00	28.107
741	54 90 81	27.221	791	62 56 81	28.125
742	55 05 64	27.240	792	62 72 64	28.142
743	55 20 49	27.258	793	62 88 49	28.160
744	55 35 36	27.276	794	63 04 36	28.178
745	55 50 25	27.295	795	63 20 25	28.196
746	55 65 16	27.313	796	63 36 16	28.213
747	55 80 09	27.331	797	63 52 09	28.231
748	55 95 04	27.350	798	63 68 04	28.249
749	56 10 01	27.368	799	63 84 01	28.267
750	56 25 00	27.386	800	64 00 00	28.284

Number	Square	Square Root	Number	Square	Square Root
801	64 16 01	28.302	851	72 42 01	29.172
802	64 32 04	28.320	852	72 59 04	29.189
803	64 4[illegible] 0[illegible]	28.337	853	72 76 09	29.206
804	64 64 16	28.355	854	72 93 16	29.223
805	64 80 25	28.373	855	73 10 25	29.240
806	64 96 36	28.390	856	73 27 36	29.257
807	65 12 49	28.408	857	73 44 49	29.275
808	65 28 64	28.425	858	73 61 64	29.292
809	65 44 81	28.443	859	73 78 81	29.309
810	65 61 00	28.460	860	73 96 00	29.326
811	65 77 21	28.478	861	74 13 21	29.343
812	65 93 44	28.496	862	74 30 44	29.360
813	66 09 69	28.513	863	74 47 69	29.377
814	66 25 96	28.531	864	74 64 96	29.394
815	66 42 25	28.548	865	74 82 25	29.411
816	66 58 56	28.566	866	74 99 56	29.428
817	66 74 89	28.583	867	75 16 89	29.445
818	66 91 24	28.601	868	75 34 24	29.462
819	67 07 61	28.618	869	75 51 61	29.479
820	67 24 00	28.636	870	75 69 00	29.496
821	67 40 41	28.653	871	75 86 41	29.513
822	67 56 84	28.671	872	76 03 84	29.530
823	67 73 29	28.688	873	76 21 29	29.547
824	67 89 76	28.705	874	76 38 76	29.563
825	68 06 25	28.723	875	76 56 25	29.580
826	68 22 76	28.740	876	76 73 76	29.597
827	68 39 29	28.758	877	76 91 29	29.614
828	68 55 84	28.775	878	77 08 84	29.631
829	68 72 41	28.792	879	77 26 41	29.648
830	68 89 00	28.810	880	77 44 00	29.665
831	69 05 61	28.827	881	77 61 61	29.682
832	69 22 24	28.844	882	77 79 24	29.698
833	69 38 89	28.862	883	77 96 89	29.715
834	69 55 56	28.879	884	78 14 56	29.732
835	69 72 25	28.896	885	78 32 25	29.749
836	69 88 96	28.914	886	78 49 96	29.766
837	70 05 69	28.931	887	78 67 69	29.783
838	70 22 44	28.948	888	78 85 44	29.799
839	70 39 21	28.965	889	79 03 21	29.816
840	70 56 00	28.983	890	79 21 00	29.833
841	70 72 81	29.000	891	79 38 81	29.850
842	70 89 64	29.017	892	79 56 64	29.866
843	71 06 49	29.034	893	79 74 49	29.883
844	71 23 36	29.052	894	79 92 36	29.900
845	71 40 25	29.069	895	80 10 25	29.916
846	71 57 16	29.086	896	80 28 16	29.933
847	71 74 09	29.103	897	80 46 09	29.950
848	71 91 04	29.120	898	80 64 04	29.967
849	72 08 01	29.138	899	80 82 01	29.983
850	72 25 00	29.155	900	81 00 00	30.000

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
901	81 18 01	30.017	951	90 44 01	30.838
902	81 36 04	30.033	952	90 63 04	30.854
903	81 54 09	30.050	953	90 82 09	30.871
904	81 72 16	30.067	954	91 01 16	30.887
905	81 90 25	30.083	955	91 20 25	30.903
906	82 08 36	30.100	956	91 39 36	30.919
907	82 26 49	30.116	957	91 58 49	30.935
908	82 44 64	30.133	958	91 77 64	30.952
909	82 62 81	30.150	959	91 96 81	30.968
910	82 81 00	30.166	960	92 16 00	30.984
911	82 99 21	30.183	961	92 35 21	31.000
912	83 17 44	30.199	962	92 54 44	31.016
913	83 35 69	30.216	963	92 73 69	31.032
914	83 53 96	30.232	964	92 92 96	31.048
915	83 72 25	30.249	965	93 12 25	31.064
916	83 90 56	30.265	966	93 31 56	31.081
917	84 08 89	30.282	967	93 50 89	31.097
918	84 27 24	30.299	968	93 70 24	31.113
919	84 45 61	30.315	969	93 89 61	31.129
920	84 64 00	30.332	970	94 09 00	31.145
921	84 82 41	30.348	971	94 28 41	31.161
922	85 00 84	30.364	972	94 47 84	31.177
923	85 19 29	30.381	973	94 67 29	31.193
924	85 37 76	30.397	974	94 86 76	31.209
925	85 56 25	30.414	975	95 06 25	31.225
926	85 74 76	30.430	976	95 25 76	31.241
927	85 93 29	30.447	977	95 45 29	31.257
928	86 11 84	20.463	978	95 64 84	31.273
929	86 30 41	30.480	979	95 84 41	31.289
930	86 49 00	30.496	980	96 04 00	31.305
931	86 67 61	30.512	981	96 23 61	31.321
932	86 86 24	30.529	982	96 43 24	31.337
933	87 04 89	30.545	983	96 62 89	31.353
934	87 23 56	30.561	984	96 82 56	31.369
935	87 42 25	30.578	985	97 02 25	31.385
936	87 60 96	30.594	986	97 21 96	31.401
937	87 79 69	30.610	987	97 41 69	31.417
938	87 98 44	30.627	988	97 61 44	31.432
939	88 17 21	30.643	989	97 81 21	31.448
940	88 36 00	30.659	990	98 01 00	31.464
941	88 54 81	30.676	991	98 20 81	31.480
942	88 73 64	30.692	992	98 40 64	31.496
943	88 92 49	30.708	993	98 60 49	31.512
944	89 11 36	30.725	994	98 80 36	31.528
945	89 30 25	30.741	995	99 00 25	31.544
946	89 49 16	30.757	996	99 20 16	31.559
947	89 68 09	30.773	997	99 40 09	31.575
948	89 87 04	30.790	998	99 60 04	31.591
949	90 06 01	30.806	999	99 80 01	31.607
950	90 25 00	30.822	1000	100 00 00	31.623

INDEX